INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
Int. J. Numer. Meth. Engng. 2000; **47**:1123–1168

Numerical implementation of a shape memory alloy thermomechanical constitutive model using return mapping algorithms

M. A. Qidwai and D. C. Lagoudas*

Aerospace Engineering Department, Center for Mechanics of Composites, Texas A & M University, College Station, TX 77843-3141, U.S.A.

SUMMARY

Previous studies by the authors and their co-workers show that the structure of equations representing shape Memory Alloy (SMA) constitutive behaviour can be very similar to those of rate-independent plasticity models. For example, the Boyd–Lagoudas polynomial hardening model has a stress-elastic strain constitutive relation that includes the transformation strain as an internal state variable, a transformation function determining the onset of phase transformation, and an evolution equation for the transformation strain. Such a structure allows techniques used in rate-independent elastoplastic behaviour to be directly applicable to SMAs. In this paper, a comprehensive study on the numerical implementation of SMA thermomechanical constitutive response using return mapping (elastic predictor-transformation corrector) algorithms is presented. The closest point projection return mapping algorithm which is an implicit scheme is given special attention together with the convex cutting plane return mapping algorithm, an explicit scheme already presented in an earlier work. The closest point algorithm involves relatively large number of tensorial operations than the cutting plane algorithm besides the evaluation of the gradient of the transformation tensor in the flow rule and the inversion of the algorithmic tangent tensor. A unified thermomechanical constitutive model, which does not take into account reorientation of martensitic variants but unifies several of the existing SMA constitutive models, is used for implementation. Remarks on numerical accuracy of both algorithms are given, and it is concluded that both algorithms are applicable for this class of SMA constitutive models and preference can only be given based on the computational cost. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: computational mechanics; return mapping algorithms; shape memory alloy (SMA); thermomechanical constitutive modelling; inelasticity

1. INTRODUCTION

Shape memory alloys (SMAs) are materials capable of changing their crystallographic structure due to changes of temperature and/or stress. These changes are referred to as martensitic phase transformations, i.e. transformation from a high symmetry austenitic phase to a low symmetry martensitic phase and *vice versa* [1]. These diffusionless phase transformations result in rate-independent but

*Correspondence to: D. C. Lagoudas, Center for Mechanics of Composites, Aerospace Engineering Department, Texas A & M University, College Station, TX 77843-3141, U.S.A.

Contract/grant sponsor: Air Force Office of Scientific Research; contract/grant number: F49620-98-1-0041

CCC 0029-5981/2000/061123–46$17.50

Received 2 December 1998
Revised 24 March 1999

fully coupled highly non-linear thermomechanical material response† with loading history dependence. This loading history dependence requires the introduction of internal state variables like the martensitic volume fraction, transformation strain, back and drag stresses due to transformation and in some cases even back and drag stresses due to transformation-induced plasticity [2]. The thermomechanical coupling is dominant in the material response even at low deformation rates and temperatures [3]. Due to these properties, SMAs exhibit the following macroscopic phenomena not present in traditional materials [1, 4]:

1. *Pseudoelasticity*: SMA is loaded above austenitic finish temperature, $A^{\circ f}$, to full transformation to the detwinned martensitic state and then unloaded fully transforming back to the austenitic state, while recovering all the transformation strain exhibiting hysteresis.
2. *Shape memory effect* (SME): SMA is loaded below austenitic start temperature, $A^{\circ s}$, while still in the austenitic phase to full transformation to the detwinned martensitic state and then unloaded fully retaining the transformation strain.

These inherent and/or induced characteristics of SMAs have given rise to their usage in a wide array of applications ranging from orthodontic wires [5] to actuators in robotic systems and self-expanding microstructures [6]. In a review paper, Birman [7] cited many studies on SMA devices used in industries which include mechanical, electronic and automotive engineering, aerospace industry, sensors, domestic appliances, and medicine. Most of this earlier work in the literature is based on the applications of SMA structural elements that are 1-D, such as wires, rods and beams. Only recently the potential of SMAs as load-bearing structural elements and actuators have been studied in the context of 3-D structures, e.g. as torque tube actuators in (smart) aircraft wings [8], active metal matrix laminates in helicopter blades [9], and actuation devices for hypersilent naval submersible [10].

These applications and the reason that even simple uniaxial loading results in a non-trivial 3-D response from SMA-based composites [11–13] demand the development of numerical tools to support the design process of SMA devices for use in load-bearing 3-D structures. Finite Element Method (FEM) is primarily used for the design and analysis of such structures and it requires the numerical implementation of the SMA constitutive model in 3-D. However, until now there was a lack of 3-D design tools due to the unavailibility of an effective 3-D SMA constitutive model based on experimental results. This is especially true when the material response is shown experimentally to be dependent on the type of deformation, i.e. tension, compression, shear [14, 15]. Only recently experiments are being conducted on SMA torque tubes under a variety of proportional and non-proportional loading paths in a combination of torsion, radial and axial deformation modes [16]. With the advent of such experimental results, it has become even more important to be able to accurately implement the 3-D SMA constitutive models using robust numerical algorithms. In literature there are several models based on 1-D experimental results that are already described or can be generalized to 3-D, e.g. the exponential model [17, 18], the cosine model [19], the polynomial model [20], the modified cosine model [21], the unified thermomechanical model [12],

†The strong thermomechanical coupling results in an apparent strain dependence of the work hardening behaviour during stress-induced phase transformation. However, this dependence is more an artifact of the prevalent thermal boundary conditions and the latent heat produced during phase transformation than an inherent material property. For example, if arrangements were made to perform austenite to martensite phase transformation of a test specimen isothermally by rapidly removing the latent heat generated regardless of the strain rate involved, then similar stress–strain material response would be obtained for all strain rates. The rate-independent model can account for this apparent rate dependence by taking into account the latent heat term in the heat equation, effectively coupling stress and temperature

etc. Since the constitutive modelling of SMAs is not the topic of interest here, the reader is referred to Fischer *et al.* [22] and Birman [7] for a thorough study of the many different approaches taken prior to 1996 from phenomenological 1-D models and their 3-D generalizations to thermodynamic-framework-based-models. For the purpose of implementation, the unified thermomechanical model is chosen here because it unifies all of the aforementioned SMA constitutive models under a consistent thermodynamic framework. It is important to note that the unified model does not take into account reorientation of martensitic variants during phase transformation.

In one of the early numerical implementations of an SMA constitutive model, Brinson and Lammering [21] studied the finite element analysis of the behaviour of shape memory alloys and their applications using 1-D truss elements. Newton's method was applied on both the weak form of momentum balance and the SMA constitutive law, and thus, replaced the materially and geometrically non-linear problem into a sequence of linear problems solved numerically. In another study, Trochu and Qian [23] performed a numerical simulation of pseudoelasticity using geometrically non-linear finite elements. A series of papers on isothermal SMA constitutive modelling and generally 1-D numerical simulations are presented by Auricchio and co-workers. Various 1-D simulations like uniaxial loading, four- and three-point bending tests are presented and compared with available experimental data [24]. A return-mapping (closest point projection) algorithm based on the backward Euler time discretization of the constitutive relations is implemented. Used in later works is the same return-mapping procedure based on two trial stresses [25, 26]. The second trial stress is used to distinguish the elastic state at the end of the phase transformation. This will be explained further in the section on the implementation of the algorithms in Appendix I. Panahandeh and Kasper [27] presented a general formulation for coupled thermomechanical simulation of shape memory alloys in the context of finite element method. They proposed a SMA constitutive model composed of a set of non-linear algebraic equations that did not require integration of a rate type evolution equation. They employed the isothermal fractional-step (staggered) method to additively decompose the coupled thermomechanical problem into a thermal and mechanical problem. Each problem is then cast into its weak form and discretized in space using the finite element method. The resultant set of algebraic equations is solved globally using a Newton–Raphson iteration scheme. Numerical simulations are performed in the context of truss finite element and its extension to multiple dimensions are shown. Govindjee and Kasper [28] numerically simulated a 1-D SMA constitutive model of Uranium–Niobium alloy accounting for plastic strains using the radial return mapping algorithm, which is a special case of the closest point projection algorithm.

Most of the studies concentrate on 1-D computational implementation of the SMA constitutive models. With regard to 3-D implementation, Reisner *et al.* [29] used a return mapping (convex cutting plane) algorithm based on the forward Euler (explicit) time discretization of the constitutive relations. A series of papers on the 3-D numerical implementation of SMA constitutive models is given by Lagoudas and co-workers. During their study of the micromechanics of active composites with SMA fibres, a SMA constitutive model was implemented using the forward Euler integration scheme [11]. In 1996, Lagoudas *et al.* presented and implemented the 3-D unified thermodynamic constitutive model based on the work of Boyd and Lagoudas [20] and they analysed the response of an active metal matrix composite with different geometric arrangements of embedded SMA fibres using the finite element method. The explicit time discretization of the transformation strain flow rule is carried out using the convex cutting plane return mapping algorithm proposed by Simo amd Ortiz [30]. Lagoudas *et al.* [13] further modelled the thermomechanical response of general active laminates with SMA strips using the layerwise finite element method based on the same

time-integration algorithm. A series of studies based on the 3-D fully coupled temperature-displacement analysis of the transformation of embedded shape memory alloy wires and ribbons was carried out by Sottos *et al.* [31] and Jonnalagadda *et al.* [32, 33]. They employed the commercially available FE software ABAQUS which solves the field equations associated with the mechanical and the heat conduction problems simultaneously using implicit integration for unconditional convergence. In a more recent series of papers, Bo and Lagoudas [2] proposed an extensive 3-D thermomechanical model based on micromechanics, where they also simulated, predicted and compared the cyclic behaviour, TWME, minor loops and material identification of 1-D SMA wires with the experiments.

It is emphasized by Moss [34] that return mapping algorithms are strain driven making them directly adaptable to displacement based finite element method (FEM). The idea of using return mapping algorithms is typical for integrating the SMA thermomechanical constitutive response. As mentioned earlier and to be shown later for the unified SMA constitutive model, rate independence and loading history dependence allow the thermomechanical constitutive behaviour to be possibly defined by a stress-elastic strain state relation that includes the transformation strain as an internal state variable, a differential evolution equation for the transformation strain and a transformation function to determine the onset of phase transformation [2, 12]. This structure implies that any proven numerical algorithm designed to integrate the rate-independent elastoplastic constitutive behaviour can be used to integrate the SMA thermomechanical elastic-transformation model.

Return mapping algorithms have been studied extensively over the years in the context of elastoplasticity for the integration of constitutive relations. They are also called elastic predictor-plastic corrector algorithms where a purely (thermo) elastic *trial state* is followed by a plastic (in this study transformation) corrector phase (return mapping). The purpose of the corrector part is to enforce continuity (iteratively) in a manner consistent with the prescribed flow rule. Return mapping algorithms may differ on the basis of the kind of discretization employed to numerically integrate the evolution equations and the (iterative) numerical procedure adopted to solve the resultant set of non-linear algebraic equations in the corrector part. Some of the return mapping algorithms proposed in the literature are the famous radial return (backward Euler based) algorithm [35] and the mean normal (mid-point rule based) algorithm [36]. A detailed review of the historical developments of the return mapping algorithms can be found in [37].

It is pointed out by Ortiz and Popov [38] that most of the return mapping algorithms employ integration rules that are particular cases of the trapezoidal and midpoint rules, suitably generalized to facilitate satisfaction of the plastic consistency condition. The stability of these integration algorithms for inelasticity is addressed by Argyris *et al.* [39] and Simo and Govindjee [40]. The following pertinent conclusions can be made from the two studies: forward marching strategies severely restricts the choice of time steps because of stability conditions; backward integration strategies improve the stability of the incremental solution; additional accuracy is obtained by midpoint integration technique and by higher order approximation of the inelastic flow rule (e.g. multistep Runge–Kutta methods).

The two integration algorithms implemented in this work are the closest point projection algorithm [41] and convex cutting plane algorithm [30, 42]. Both algorithms rely on the strain-driven nature of the problem in the sense that state variables are computed for a given deformation history. The difference arises in the corrector (or relaxation) part of the problem. The application of the closest point projection algorithm results in a set of non-linear algebraic equations solved using Newton's iteration method. The closest point algorithm is unconditionally stable provided the yield

surface is convex, and it is first-order accurate [41].‡ On the other hand, the convex cutting plane method is based on the explicit integration of the differential equations and its advantage lies in its simplicity and less computations.

At this point, the outline of the paper is given as follows. In Section 2, the unified SMA thermomechanical constitutive model, which does not take into account reorientation of martensitic variants, is briefly discussed and the derivation of the continuum tangent moduli tensors, which are required for use in the FEM if the convex cutting plane algorithm is implemented, is given. The general return mapping algorithm is explained in the context of SMAs in Section 3 followed by the incremental derivation of the two algorithms. Benchmark cases, like isothermal pseudoelasticity and constant stress thermally induced phase transformation, are analysed using FEM to study the convergence characteristics of the algorithms for the unified SMA model and results are given in Section 4. In the same section, the numerical accuracy of the algorithms is studied by simulating SMA torque tube boundary value problems under varying strain increment sizes, and a simple error estimation based on the finite element solution is provided. At the end, SMA torque tubes are simulated under different loading paths in stress space to study the path dependence of the material response. In Appendix I, the loading/unloading conditions in stress space and the issue of the martensitic volume fraction becoming greater than one or less than zero during transformation correction are discussed, and the derivation of the consistent tangent moduli tensors§ for the closest point projection algorithm is given.

2. SMA CONSTITUTIVE MODEL

For study purposes, the unified SMA constitutive model proposed by Lagoudas *et al.* [12] based on the thermodynamic framework of Boyd and Lagoudas [20] is chosen. The advantage of using this model is that a broad class of SMA constitutive models can be formulated under its framework; therefore, the present results are applicable for all of them. A brief overview of the model is given here and the reader is referred to the aforementioned works for a detailed study.

The total Gibbs free energy, G, of a polycrystalline SMA, assuming linear thermoelastic response and non-linear transformation-hardening behaviour, is given by

$$G(\boldsymbol{\sigma}, T, \xi, \boldsymbol{\varepsilon}^t) := -\frac{1}{2}\frac{1}{\rho}\boldsymbol{\sigma} : \mathcal{S} : \boldsymbol{\sigma} - \frac{1}{\rho}\boldsymbol{\sigma} : [\boldsymbol{\alpha}(T - T_0) + \boldsymbol{\varepsilon}^t] + c\left[(T - T_0) - T\ln\left(\frac{T}{T_0}\right)\right] - s_0 T + u_0 + f(\xi) \tag{1}$$

where $\boldsymbol{\sigma}, \boldsymbol{\varepsilon}^t, \xi, T$ and T_0 are the Cauchy stress tensor, transformation strain tensor, martensitic volume fraction, temperature and reference temperature, respectively. The material constants $\mathcal{S}, \boldsymbol{\alpha}, \rho, c$, s_0 and u_0 are the effective compliance tensor, effective thermal expansion coefficient tensor, density, effective specific heat, effective specific entropy at the reference state, and effective specific internal energy at the reference state, respectively. All the above effective material properties are

‡First-order accuracy or consistency is defined as the condition that the numerically integrated variables agree with their exact values to within second-order terms in the step size, Δt

§The concept of consistent tangent moduli tensors is introduced by Simo and Taylor [43]

defined in terms of the internal state variable, ξ, by the rule of mixtures as

$$\begin{aligned} &\mathcal{S} := \mathcal{S}^{\mathrm{A}} + \xi(\mathcal{S}^{\mathrm{M}} - \mathcal{S}^{\mathrm{A}}), \quad \boldsymbol{\alpha} := \boldsymbol{\alpha}^{\mathrm{A}} + \xi(\boldsymbol{\alpha}^{\mathrm{M}-\mathrm{A}}) \\ &c := c^{\mathrm{A}} + \xi(c^{\mathrm{M}} - c^{\mathrm{A}}), \quad s_0 := s_0^{\mathrm{A}} + \xi(s_0^{\mathrm{M}} - s_0^{\mathrm{A}}) \quad u_0 := u_0^{\mathrm{A}} + \xi(u_0^{\mathrm{M}} - u_0^{\mathrm{A}}). \end{aligned} \tag{2}$$

The function, $f(\xi)$, in equation (1) accounts for transformation hardening due to phase interaction.

The Gibbs free energy is related to the internal energy, u, by the following Legendre transformation:

$$u(\boldsymbol{\sigma}, s, \xi, \boldsymbol{\varepsilon}^t) = G + Ts + \frac{1}{\rho}\boldsymbol{\sigma} : \boldsymbol{\varepsilon} \tag{3}$$

where $\boldsymbol{\varepsilon}$ is the total strain tensor. Following the formalism of Truesdell and Noll [44], G and u are substituted in the first and second law of thermodynamics to derive thermodynamic constraints on state of the system resulting in the following constitutive relations:

$$\boldsymbol{\varepsilon} = -\rho\partial_{\boldsymbol{\sigma}} G = \mathcal{S} : \boldsymbol{\sigma} + \boldsymbol{\alpha}(T - T_0) + \boldsymbol{\varepsilon}^t \tag{4}$$

$$s = -\partial_T G = \frac{1}{\rho}\boldsymbol{\sigma} : \boldsymbol{\alpha} + c \ln\left(\frac{T}{T_0}\right) + s_0 \tag{5}$$

Moreover, the local dissipation rate is given by

$$\boldsymbol{\sigma} : \dot{\boldsymbol{\varepsilon}}^t - \rho\partial_\xi G\dot{\xi} \geqslant 0. \tag{6}$$

To simplify the analysis for the case of phase transformation without reorientation, it is assumed that any change in the current state of the system is only possible by a change in the martensitic volume fraction and that any other internal state variable evolves with it [2]. Based on this assumption, the following relation is introduced to relate the evolution of transformation strain tensor to the evolution of the martensitic volume fraction:

$$\dot{\boldsymbol{\varepsilon}}^t = \boldsymbol{\Lambda}\dot{\xi} \tag{7}$$

where $\boldsymbol{\Lambda}$ is the transformation tensor which determines the transformation strain (flow) direction and is assumed to have the following form:

$$\boldsymbol{\Lambda} := \begin{cases} \frac{3}{2}H\frac{\boldsymbol{\sigma}'}{\bar{\sigma}}, & \dot{\xi} > 0 \\ H\frac{\boldsymbol{\varepsilon}^{t-r}}{\bar{\varepsilon}^{t-r}}, & \dot{\xi} < 0 \end{cases} \tag{8}$$

where H is the maximum uniaxial transformation strain, $\boldsymbol{\varepsilon}^{t-r}$ is the transformation strain at the reversal of phase transformation, and

$$\bar{\sigma} = \sqrt{\tfrac{3}{2}}\|\boldsymbol{\sigma}'\|, \quad \boldsymbol{\sigma}' = \boldsymbol{\sigma} - \tfrac{1}{3}(\mathrm{tr}\,\boldsymbol{\sigma})\mathbf{1}, \quad \bar{\varepsilon}^{t-r} = \sqrt{\tfrac{2}{3}}\|\boldsymbol{\varepsilon}^{t-r}\|, \tag{9}$$

where $\|\cdot\|^2 = (\cdot : \cdot)$ is the inner product of the enclosed quantity. It is important to note here that the transformation strain rate, $\dot{\boldsymbol{\varepsilon}}^t$, is only a part of the total inelastic strain rate, $\dot{\boldsymbol{\varepsilon}}^{\mathrm{in}}$.

Using equation (7), the total dissipation rate can be rewritten as

$$(\boldsymbol{\sigma} : \boldsymbol{\Lambda} - \rho\partial_\xi G)\dot{\xi} = \pi\dot{\xi} \geqslant 0 \tag{10}$$

where π is the thermodynamic force conjugate to ξ, which is obtained after substitution from equation (1) as

$$\pi = \boldsymbol{\sigma} : \boldsymbol{\Lambda} + \frac{1}{2}\boldsymbol{\sigma} : \Delta\mathcal{S} : \boldsymbol{\sigma} + \Delta\boldsymbol{\alpha} : \boldsymbol{\sigma}(T - T_0) + \rho\Delta c \left[(T - T_0) - T \ln\left(\frac{T}{T_0}\right)\right] + \rho\Delta s_0 T - \frac{\partial f}{\partial \xi} - \rho\Delta u_0 \tag{11}$$

The terms that are defined with the prefix Δ in equation (11) indicate the difference of a quantity between the martensitic and austenitic phases, and they are given by

$$\begin{aligned} &\Delta\mathcal{S} = \mathcal{S}^{\mathrm{M}} - \mathcal{S}^{\mathrm{A}}, \quad \Delta\boldsymbol{\alpha} = \boldsymbol{\alpha}^{\mathrm{M}} - \boldsymbol{\alpha}^{\mathrm{A}}, \quad \Delta c = c^{\mathrm{M}} - c^{\mathrm{A}} \\ &\Delta s_0 = s_0^{\mathrm{M}} - s_0^{\mathrm{A}}, \quad \Delta u_0 = u_0^{\mathrm{M}} - u_0^{\mathrm{A}} \end{aligned} \tag{12}$$

The transformation function is defined in terms of the thermodynamic force, π, as follows:

$$\Phi := \begin{cases} \pi - Y^*, & \dot{\xi} > 0 \\ -\pi - Y^*, & \dot{\xi} < 0 \end{cases} \tag{13}$$

where Y^* is the measure of internal dissipation due to phase transformation. Constraints on the evolution of the martensitic volume fraction, derived from a maximum transformation dissipation principle similar to the maximum plastic dissipation principle, are expressed in terms of the Kuhn–Tucker inequalities as

$$\begin{aligned} &\dot{\xi} \geqslant 0, \quad \Phi(\boldsymbol{\sigma}, T, \xi) \leqslant 0, \quad \Phi\dot{\xi} = 0 \\ &\dot{\xi} \leqslant 0, \quad \Phi(\boldsymbol{\sigma}, T, \xi) \leqslant 0, \quad \Phi\dot{\xi} = 0 \end{aligned} \tag{14}$$

Here, $\Phi(\boldsymbol{\sigma}, T, \xi)$ signifies the transformation function that defines the elastic domain. The inequality constraints on $\Phi(\boldsymbol{\sigma}, T, \xi)$ may also be called as the transformation (consistency) condition and viewed as a constraint on the state variables' admissibility. Along any loading path, all the relations should hold simultaneously. For $\Phi < 0$, equation (14) requires $\dot{\xi} = 0$ and elastic response is obtained. On the other hand, the forward-phase transformation (austenite to martensite) is characterized by $\Phi = 0$ and $\dot{\xi} > 0$, while the reverse-phase transformation (martensite to austenite) is characterized by $\Phi = 0$ and $\dot{\xi} < 0$.

The transformation function is plotted for $\xi = 0$ and $\xi = 1$ in Figures 1 and 2 in stress-temperature space, and in Figures 3(a) and 3(b) in stress space only, for both forward and reverse transformations at the austenitic finish temperature. In Figures 2 and 3(b), the orientation of the transformation function in the reverse direction is defined by the transformation tensor, $\boldsymbol{\Lambda}$, which is history dependent in the sense that it is defined by the previous forward transformation. In the present case, it is defined by the transformation strain tensor at the end of the full forward-phase transformation under isothermal uniaxial loading conditions (see (8)). It can be seen from Figures 1 and 2 that the transformation function represents a transformation surface for a given set of internal state variables and the two surfaces for $\xi = 0$ and $\xi = 1$ in each figure are the upper and lower bound of transformation surfaces. Any stress state not on or between these transformation surfaces is elastic. For example, if SMA is loaded isothermally in the austenitic state ($\xi = 0$), it behaves elastically, then transforms fully into martensite ($\xi = 1$) and then again exhibits elastic behaviour. Also notice in Figures 1 and 3(a) that this model assumes symmetry (von-Mises type) between the tensile and compressive behaviours for forward transformation. However, recent experimental results show

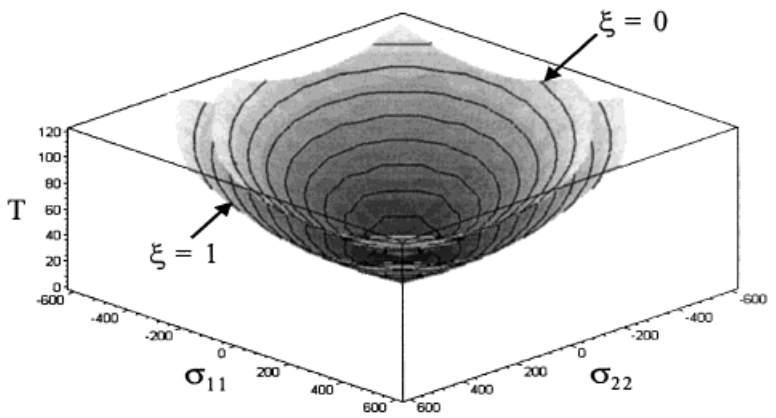


Figure 1. Plot of transformation function in σ_{11}–σ_{22}–T space for forward transformation given the SMA material properties in Table I at A^{of}

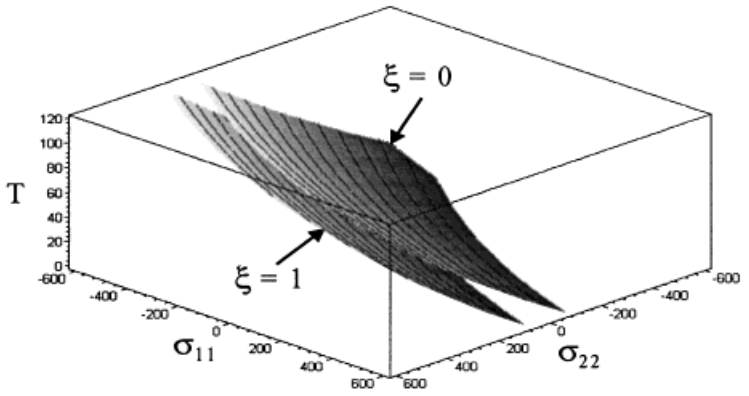


Figure 2. Plot of transformation function in σ_{11}–σ_{22}–T space for reverse transformation given the SMA material properties in Table I at A^{of}

asymmetry between the two behaviours. This effect can also be conveniently captured in this model by the introduction of either a first stress-invariant or a third deviatoric stress invariant term, either directly into the Gibbs free energy expression (equation (1)) or into the transformation function (equation (13)) as done by Gillet *et al.* [45].

The hardening function, $f(\xi)$, is responsible for the transformation-induced strain hardening in the SMA material and is given by

$$f(\xi) = \begin{cases} \frac{1}{2}\rho b^{M}\xi^2 + (\mu_1 + \mu_2)\xi, & \dot{\xi} > 0 \\ \frac{1}{2}\rho b^{A}\xi^2 + (\mu_1 - \mu_2)\xi, & \dot{\xi} < 0 \end{cases} \tag{15}$$

where $\rho b^{M}, \rho b^{A}, \mu_1$ and μ_2 are transformation strain hardening material constants. The above choice of quadratic dependence of the Gibbs free energy, G, on the martensitic volume fraction, ξ, can be

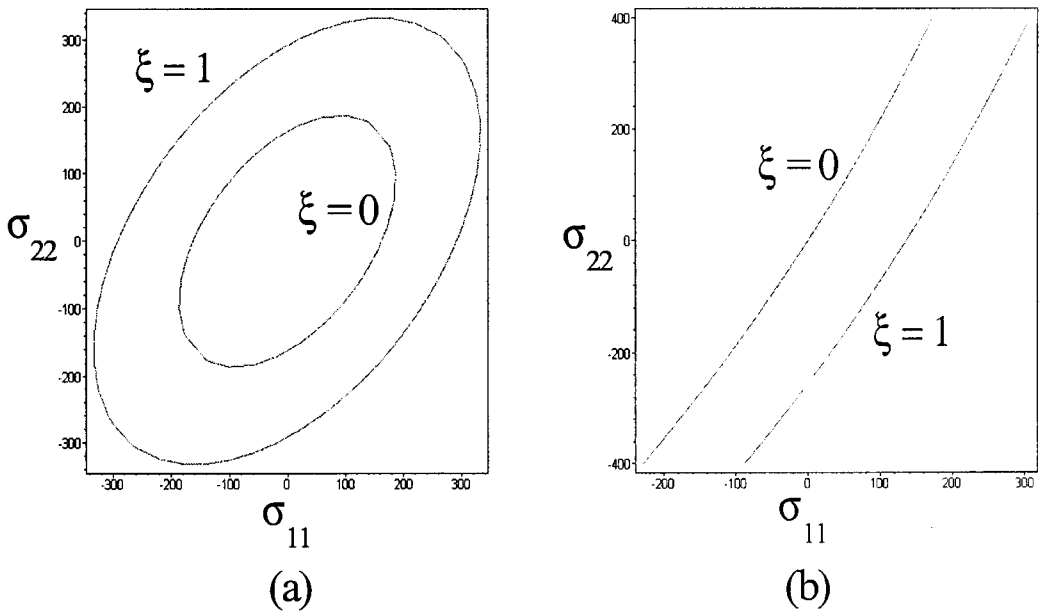


Figure 3. Plot of transformation function in $\sigma_{11}-\sigma_{22}$ space for: (a) forward transformation; (b) reverse transformation given the SMA material properties in Table I at $A^{\circ f}$

replaced by trigonometric [19, 21], logarithmic [17], or any other function to accommodate other constitutive models.

The normality of the total inelastic strain rate can be shown by the following derivation. Starting with the expression for the total strain rate using chain rule

$$\dot{\boldsymbol{\varepsilon}}(\boldsymbol{\sigma}, T, \xi) = \partial_{\boldsymbol{\sigma}}\boldsymbol{\varepsilon} : \dot{\boldsymbol{\sigma}} + \partial_T \boldsymbol{\varepsilon}\dot{T} + \partial_{\xi}\boldsymbol{\varepsilon}\dot{\xi} \tag{16}$$

the total inelastic strain rate can be defined in terms of the change of total strain with respect to the martensitic volume fraction, i.e.

$$\dot{\boldsymbol{\varepsilon}}^{\text{in}} := \partial_{\xi}\boldsymbol{\varepsilon}\dot{\xi} \tag{17}$$

The right-hand side in the above expression is obtained by taking the partial derivative of the strain–stress constitutive relation (equation (4)) with respect to ξ, i.e.

$$\dot{\boldsymbol{\varepsilon}}^{\text{in}} = [\Delta\mathcal{S} : \boldsymbol{\sigma} + \Delta\boldsymbol{\alpha}(T - T_0) + \partial_{\xi}\boldsymbol{\varepsilon}^t]\dot{\xi} \tag{18}$$

The first two terms in the above equation are the contribution to the inelastic strain rate from the change of material properties with respect to the martensitic volume fraction. Using the transformation strain flow rule to define the last term in equation (18) results in

$$\dot{\boldsymbol{\varepsilon}}^{\text{in}} = [\Delta\mathcal{S} : \boldsymbol{\sigma} + \Delta\boldsymbol{\alpha}(T - T_0) + \boldsymbol{\Lambda}]\dot{\xi} \tag{19}$$

However, it can be easily shown by substituting equation (11) into equation (13) and taking its derivative with respect to the stress tensor, $\boldsymbol{\sigma}$, that

$$\partial_{\boldsymbol{\sigma}}\Phi = \begin{cases} \Delta\mathcal{S}:\boldsymbol{\sigma} + \Delta\boldsymbol{\alpha}(T - T_0) + \boldsymbol{\Lambda}, & \dot{\xi} > 0 \\ -\Delta\mathcal{S}:\boldsymbol{\sigma} - \Delta\boldsymbol{\alpha}(T - T_0) - \boldsymbol{\Lambda}, & \dot{\xi} < 0 \end{cases} \tag{20}$$

and, therefore,

$$\dot{\boldsymbol{\varepsilon}}^{\text{in}} = \begin{cases} \dot{\xi}\partial_{\boldsymbol{\sigma}}\Phi, & \dot{\xi} > 0 \\ -\dot{\xi}\partial_{\boldsymbol{\sigma}}\Phi, & \dot{\xi} < 0 \end{cases} \tag{21}$$

Equation (21) shows the flow rule for the inelastic strain evolution to be associative in the essence of J^2 plasticity. It is noticed that $\dot{\xi}$ acts like a transformation parameter in analogy to the plastic parameter in associative plasticity. The above equation also implies that the transformation strain rate does not follow the normality rule due to the change in material properties with transformation.

2.1. Continuum tangent moduli tensors

In an incremental displacement-based finite element analysis, the increment of the stress tensor is obtained through the implementation of the incremental SMA constitutive model for given increments of strain and temperature. The SMA constitutive model can be written in an incremental form as follows:

$$\mathrm{d}\boldsymbol{\sigma} = \mathscr{L}:\mathrm{d}\boldsymbol{\varepsilon} + \boldsymbol{\Theta}\ \mathrm{d}T \tag{22}$$

where $\mathscr{L}$ is the tangent stiffness tensor and $\boldsymbol{\Theta}$ is the tangent thermal moduli tensor. These two tensors are needed for the execution of the global finite element analysis (FEA).

In order to derive $\mathscr{L}$ and $\boldsymbol{\Theta}$, the constitutive relation in equation (4) is written in differential form and the transformation strain flow rule (equation (7)) is substituted to get

$$\mathrm{d}\boldsymbol{\sigma} = \mathcal{S}^{-1}:\{\mathrm{d}\boldsymbol{\varepsilon} - \boldsymbol{\alpha}\mathrm{d}T - [\Delta\mathcal{S}:\boldsymbol{\sigma} + \Delta\boldsymbol{\alpha}(T - T_0) + \boldsymbol{\Lambda}]\mathrm{d}\xi\} \tag{23}$$

Using equation (20), the above equation reduces to

$$\mathrm{d}\boldsymbol{\sigma} = \mathcal{S}^{-1}:\left(\mathrm{d}\boldsymbol{\varepsilon} - \boldsymbol{\alpha}\mathrm{d}T - \mathrm{d}\xi\begin{cases} \partial_{\boldsymbol{\sigma}}\Phi, & \dot{\xi} > 0 \\ -\partial_{\boldsymbol{\sigma}}\Phi, & \dot{\xi} < 0 \end{cases}\right) \tag{24}$$

On the other hand, the differentiation of the transformation function (consistency condition) results in

$$\mathrm{d}\Phi = \partial_{\boldsymbol{\sigma}}\Phi:\mathrm{d}\boldsymbol{\sigma} + \partial_T\Phi\mathrm{d}T + \partial_\xi\Phi\mathrm{d}\xi = 0 \tag{25}$$

For forward transformation ($\dot{\xi} > 0$), an expression for the differential of the martensitic volume fraction, $\mathrm{d}\xi$, is obtained by substituting $\mathrm{d}\boldsymbol{\sigma}$ from equation (24a) into equation (25), i.e.

$$\mathrm{d}\xi = \frac{\partial_{\boldsymbol{\sigma}}\Phi:\mathcal{S}^{-1}:\mathrm{d}\boldsymbol{\varepsilon} + (\partial_T\Phi - \partial_{\boldsymbol{\sigma}}\Phi:\mathcal{S}^{-1}:\boldsymbol{\alpha})\mathrm{d}T}{\partial_{\boldsymbol{\sigma}}\Phi:\mathcal{S}^{-1}:\partial_{\boldsymbol{\sigma}}\Phi - \partial_\xi\Phi} \tag{26}$$

Now equation (26) can be used to eliminate $\mathrm{d}\xi$ in equation (24a) and obtain the relationship between the stress, strain and temperature increments as

$$\mathrm{d}\boldsymbol{\sigma} = \left[\mathcal{S}^{-1} - \frac{\mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi \otimes \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi}{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi - \partial_{\xi}\Phi}\right] : \mathrm{d}\varepsilon + \mathcal{S}^{-1} : \left[\partial_{\boldsymbol{\sigma}}\Phi \left(\frac{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \boldsymbol{\alpha} - \partial_T\Phi}{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi - \partial_{\xi}\Phi}\right) - \boldsymbol{\alpha}\right] \mathrm{d}T \tag{27}$$

The tangent stiffness tensor, $\mathscr{L}$, and tangent thermal moduli tensor, $\boldsymbol{\Theta}$, are then given, after some simplifications of equation (27) by

$$\mathscr{L} = \mathcal{S}^{-1} - \mathbf{A} \otimes \mathbf{A} \tag{28}$$

$$\boldsymbol{\Theta} = -\mathscr{L} : \boldsymbol{\alpha} - \frac{\partial_T\Phi}{a}\mathbf{A} \tag{29}$$

where

$$a := \sqrt{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi - \partial_{\xi}\Phi}$$
$$\mathbf{A} := \frac{\mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi}{\sqrt{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi - \partial_{\xi}\Phi}} = \frac{\mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi}{a} \tag{30}$$

Following the same methodology for the reverse-phase transformation, i.e. substituting equation (24b) into equation (25) gives

$$\mathrm{d}\xi = -\frac{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \mathrm{d}\boldsymbol{\varepsilon} + (\partial_T\Phi - \partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \boldsymbol{\alpha})\mathrm{d}T}{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi + \partial_{\xi}\Phi} \tag{31}$$

Using equations (24b) and (31), the differential of stress can be obtained in terms of $\mathrm{d}\boldsymbol{\varepsilon}$ and $\mathrm{d}T$ in a similar way as described for the forward transformation, and $\mathscr{L}$ and $\boldsymbol{\Theta}$ are then given by

$$\mathscr{L} = \mathcal{S}^{-1} - \mathbf{B} \otimes \mathbf{B} \tag{32}$$

$$\Theta = -\mathscr{L} : \boldsymbol{\alpha} - \frac{\partial_T\Phi}{b}\mathbf{B} \tag{33}$$

where

$$b := \sqrt{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi + \partial_{\xi}\Phi}$$
$$\mathbf{B} := \frac{\mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi}{\sqrt{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi + \partial_{\xi}\Phi}} = \frac{\mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi}{b} \tag{34}$$

3. RETURN MAPPING ALGORITHMS

For the system of algebraic and ordinary differential equations (ODEs) given by equations (4) and (7), and constrained by equation (14), it is assumed that the strain and temperature history $\boldsymbol{\varepsilon}(t)$ and $T(t)$ is known for $t \in [0, \hat{t}]$, as well as the initial conditions for $\boldsymbol{\varepsilon}^t$ and ξ at $t = 0$. Then the

evolution equations for the transformation strain are discretized using the generalized trapezoidal rule as

$$\boldsymbol{\varepsilon}^t_{n+1} = \boldsymbol{\varepsilon}^t_n + (\xi_{n+1} - \xi_n)[(1-\beta)\boldsymbol{\Lambda}_n + \beta\boldsymbol{\Lambda}_{n+1}] \tag{35}$$

where β ranges from $[0,1]$ and subscript n indicates function evaluations at time t_n (similarly for $n+1$), assuming time increment $\Delta t = t_{n+1} - t_n$. For different values of β, various difference operators can be obtained. In the present formulation, two such values are chosen which together demonstrate all the characteristics of this family of integration rules. In particular, for $\beta = 1$ the implicit (backward) Euler integration rule is obtained. This implicit algorithm is implemented for the set of equations and Kuhn–Tucker conditions given by equations (4), (7) and (14) in Section 3.2, while in Section 3.3 the explicit (forward) Euler integration rule ($\beta = 0$) is implemented. Both integration methods result in a set of non-linear algebraic equations whose implicit nature require multiple iterations to obtain a solution.

The return mapping algorithm divides the problem posed by this set of non-linear algebraic equations in an additive split in such a way that requires iterations only for the second part containing the transformation behaviour.

3.1. A general view of thermoelastic prediction-transformation correction return mapping

The general thermoelastic-transformation problem of evolution (equations (4), (7) and (14)) can be split into two problems using an additive split as shown below

1. A thermoelastic predictor problem that is described by assuming that the increment of the transformation strain is zero

$$\begin{aligned} \dot{\boldsymbol{\varepsilon}} &= \dot{\boldsymbol{\varepsilon}}(t) \\ \dot{T} &= \dot{T}(t) \\ \dot{\boldsymbol{\varepsilon}}^t &= 0 \end{aligned} \tag{36}$$

2. The transformation corrector problem that restores consistency if the predicted thermoelastic state is outside the transformation surface ($\Phi > 0$) given by

$$\begin{aligned} \dot{\boldsymbol{\varepsilon}} &= 0 \\ \dot{T} &= 0 \\ \dot{\boldsymbol{\varepsilon}}^t &= \boldsymbol{\Lambda}\dot{\xi} \end{aligned} \tag{37}$$

with initial conditions provided by the solution of the elastic predictor problem, and the increment $\dot{\xi}$ is to be found by satisfying the transformation consistency condition.

Before discretizing the transformation correction problem in equation (37) according to a particular integration rule, a brief derivation is performed here on how this correction takes place regardless of the chosen integration procedure. The details of the numerical integration in the correction phase for two algorithms will follow in successive sections. Recalling that the history of the total strain tensor, $\boldsymbol{\varepsilon}(t)$, and temperature, $T(t)$, are given, the transformation function during correction will be varying only with respect to changes in $\boldsymbol{\sigma}$ and ξ. Since $\mathrm{d}\boldsymbol{\varepsilon}$ and $\mathrm{d}T$ are zero

during transformation correction, equation (24) results in

$$\frac{\mathrm{d}\boldsymbol{\sigma}}{\mathrm{d}\xi} = \begin{cases} -\mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi, & \dot{\xi} > 0 \\ \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi, & \dot{\xi} < 0 \end{cases} \tag{38}$$

Note that the evolution of stress state towards the transformation surface starts at the trial thermoelastic state and follows the steepest descent path with respect to the transformation surface, Φ, defined in the metric of elasticities [46]. Transformation consistency is enforced by determining the intersection of the stress evolution curve with the boundary of the transformation function (return mapping).

Further understanding of the transformation correction problem can be obtained by examining the evolution of Φ from the value it takes in the thermoelastic prediction to the final value of zero. Differentiating Φ with respect to ξ, recalling that T is kept constant during transformation correction, and using the expression for $\mathrm{d}\boldsymbol{\sigma}/\mathrm{d}\xi$ in equation (38), we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\xi}\Phi(\boldsymbol{\sigma}, T|_{\text{fixed}}, \xi) &= \partial_{\boldsymbol{\sigma}}\Phi : \frac{\mathrm{d}\boldsymbol{\sigma}}{\mathrm{d}\xi} + \partial_{\xi}\Phi \\ &= \begin{cases} -\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : & \partial_{\boldsymbol{\sigma}}\Phi + \partial_{\xi}\Phi, \quad \dot{\xi} > 0 \\ \partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi + \partial_{\xi}\Phi, & \dot{\xi} < 0 \end{cases} \\ &= \begin{cases} -\|\partial_{\boldsymbol{\sigma}}\Phi\|^2_{\mathcal{S}^{-1}} - \rho b^{\mathrm{M}} < 0, & \dot{\xi} > 0 \\ \|\partial_{\boldsymbol{\sigma}}\Phi\|^2_{\mathcal{S}^{-1}} + \rho b^{\mathrm{A}} > 0, & \dot{\xi} < 0 \end{cases} \end{aligned} \tag{39}$$

where $\partial_{\xi}\Phi$ is derived using equations (11) and (13), and $\|\partial_{\boldsymbol{\sigma}}\Phi\|^2_{\mathcal{S}^{-1}} = \sqrt{\partial_{\boldsymbol{\sigma}}\Phi : \mathcal{S}^{-1} : \partial_{\boldsymbol{\sigma}}\Phi}$ is the norm of $\partial_{\boldsymbol{\sigma}}\Phi$ induced by the elasticity tensor, $\mathcal{S}^{-1}$. Equation (39a) implies that $\Phi(\boldsymbol{\sigma}, \xi)$ monotonically decreases with increasing ξ in the forward transformation with a negative slope. Additionally, the function Φ is convex for forward transformation as shown in Figure 4. These characteristics make the transformation correction well suited for a solution by Newton's method, which is used in this work. Similarly, equation (39b) implies that $\Phi(\boldsymbol{\sigma}, \xi)$ vs. ξ monotonically decreases with decreasing ξ with a positive slope. However, the function Φ is concave for reverse transformation as shown in Figure 5. These two observations imply that the application of Newton's method for the transformation correction will only be locally convergent.

Attention is now focused on the application of the return mapping algorithm on the discretized SMA constitutive model based on the backward Euler and forward Euler integration methods. Again consider the time interval $[0, \hat{t}]$ and assume that at time $t_n \in [0, \hat{t}]$, the total strain, the temperature, and the internal state variables, which are the transformation strain and the martensitic volume fraction, are known, i.e.

$$\boldsymbol{\varepsilon}|_{t=t_n} = \boldsymbol{\varepsilon}_n, \quad T|_{t=t_n} = T_n, \quad \boldsymbol{\varepsilon}^t|_{t=t_n} = \boldsymbol{\varepsilon}^t_n, \quad \xi|_{t=t_n} = \xi_n \tag{40}$$

Given the increments of strain and temperature field, $\{\Delta\boldsymbol{\varepsilon}_{n+1} = \boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\varepsilon}_n, \Delta T = T_{n+1} - T_n\}$, where $\boldsymbol{\varepsilon}_{n+1} = \boldsymbol{\varepsilon}(t_{n+1})$, $T_{n+1} = T(t_{n+1})$ at time $t_{n+1} \in [0, \hat{t}]$, $t_{n+1} > t_n$, the task is to update the field variables $\{\boldsymbol{\sigma}, \boldsymbol{\varepsilon}^t, \xi\}$ to t_{n+1}, in a way consistent with the constitutive relations developed in Section 2.

3.2. Closest point projection return mapping algorithm

The main idea in this algorithm is to integrate the transformation flow rule for the transformation correction using the backward Euler method ($\beta = 1$ in equation (35)) resulting in a non-linear

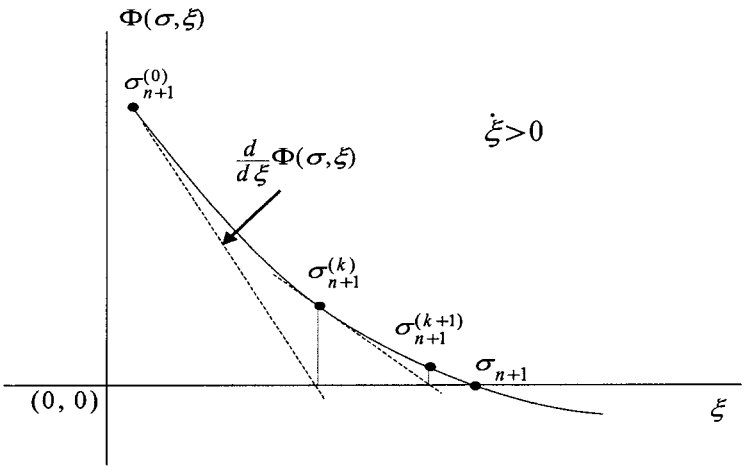


Figure 4. Φ monotonically decreases with increasing ξ with a negative slope for the forward transformation making it possible to use the Newton's method to solve for its zero as a function of ξ during transformation correction

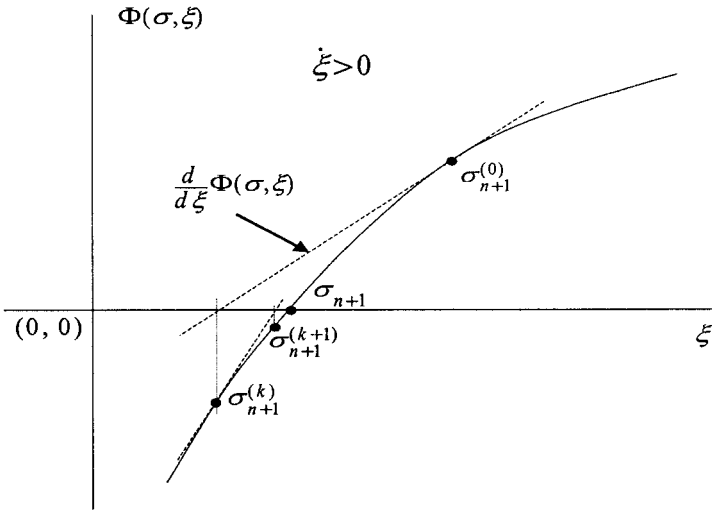


Figure 5. Φ monotonically decreases with decreasing ξ with a positive slope for the reverse transformation making it possible to use the Newton's method to solve for its zero as a function of ξ during transformation correction as shown

algebraic set of equations, which are solved using Newton iteration method. The numerical discretization with the initial conditions given in equation (40) is performed on the SMA constitutive model (equations (4), (7), (14)) as

$$\boldsymbol{\sigma}_{n+1} = \mathcal{S}^{-1}_{n+1} : [\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_{n+1}(T_{n+1} - T_0) - \boldsymbol{\varepsilon}^t_{n+1}] \tag{41}$$

$$\boldsymbol{\varepsilon}^t_{n+1} = \boldsymbol{\varepsilon}^t_n + (\xi_{n+1} - \xi_n)\boldsymbol{\Lambda}_{n+1} \tag{42}$$

constrained by the discrete Kuhn–Tucker optimality conditions

$$\begin{aligned}
&(\xi_{n+1}-\xi_n)>0, \quad \Phi_{n+1}(\boldsymbol{\sigma}_{n+1},T_{n+1},\xi_{n+1})\leqslant 0, \quad (\xi_{n+1}-\xi_n)\Phi_{n+1}(\boldsymbol{\sigma}_{n+1},T_{n+1},\xi_{n+1})=0\\
&(\xi_{n+1}-\xi_n)<0, \quad \Phi_{n+1}(\boldsymbol{\sigma}_{n+1},T_{n+1},\xi_{n+1})\leqslant 0, \quad (\xi_{n+1}-\xi_n)\Phi_{n+1}(\boldsymbol{\sigma}_{n+1},T_{n+1},\xi_{n+1})=0
\end{aligned} \tag{43}$$

3.2.1. Thermoelastic prediction. The thermoelastic prediction problem is given by

$$\begin{aligned}
\boldsymbol{\varepsilon}_{n+1} &= \boldsymbol{\varepsilon}_n + \Delta\boldsymbol{\varepsilon}_{n+1}\\
T_{n+1} &= T_n + \Delta T_{n+1}\\
\boldsymbol{\varepsilon}^{t\,(0)}_{n+1} &= \boldsymbol{\varepsilon}^t_n\\
\xi^{(0)}_{n+1} &= \xi_n
\end{aligned} \tag{44}$$

where $\{\Delta T_{n+1}, \Delta\boldsymbol{\varepsilon}_{n+1}\}$ are the specified strain tensor and temperature increments, respectively, over the time step $[t_n, t_{n+1}]$, and $(\cdot)^{(0)}$ denotes the values obtained in the prediction stage. The above set of equations (equations (41)–(43)) can be reformulated in a form suitable for the computational implementation (operator split) by introducing the following trial thermoelastic state:

$$\boldsymbol{\sigma}^{(0)}_{n+1} := \mathcal{S}^{-1}_n : [\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_n(T_{n+1}-T_0) - \boldsymbol{\varepsilon}^t_n] \tag{45}$$

$$\Phi^{(0)}_{n+1} := \Phi[\boldsymbol{\sigma}^{(0)}_{n+1}, T_{n+1}, \xi_n] \tag{46}$$

where the material properties $\mathcal{S}$ and $\boldsymbol{\alpha}$ are based on ξ_n. Note that this problem is solved using the converged internal state variable of the previous time step and the increment of total strain tensor. Physically the reformulation means that this trial state is obtained by neglecting the transformation during the time step, i.e. $\Delta\xi_{n+1}=0$. At this stage, if the transformation criterion is satisfied, that is, $\Phi(\boldsymbol{\sigma}^{(0)}_{n+1}, T_{n+1}, \xi_n)\leqslant 0$ then this trial thermoelastic state is the final state. However, if $\Phi(\boldsymbol{\sigma}^{(0)}_{n+1}, T_{n+1}, \xi_n)>0$, then the Kuhn–Tucker conditions are violated and the trial state lies outside the transformation surface. Transformation correction employing backward Euler integration of the flow rule is then used to restore consistency and the solution $[\boldsymbol{\sigma}^{(0)}_{n+1}, T_{n+1}, \xi_n]$ is taken as the initial condition for the transformation corrector phase.

3.2.2. Transformation correction. If the trial stress, $\boldsymbol{\sigma}^{(0)}_{n+1}$ and the converged internal state variables, $\boldsymbol{\varepsilon}^t_n$ and ξ_n, violates the transformation conditions in equation (43), i.e. $\Phi^{(0)}_{n+1}>0$, the following transformation correction procedure is adopted at the $(n+1)$th loading increment for the solution of equations (41)–(42) with initial guess equation (45) and constraints given by equations (43). It is important to note here that replacing $\boldsymbol{\sigma}_n$ by $\boldsymbol{\sigma}^{(0)}_{n+1}$ implies that the implicit integration of the transformation strain flow rule is carried from the initial thermoelastic prediction to the final state.

The non-linear system of these algebraic equations is solved by defining the transformation condition valid for the transformation corrector phase and transformation strain residual based on implicit backward Euler integration from equation (42) for the kth iteration as follows:

$$\Phi^{(k)}_{n+1} := \Phi(\boldsymbol{\sigma}^{(k)}_{n+1}, \xi^{(k)}_{n+1}) \tag{47}$$

$$\mathbf{R}^{t(k)}_{n+1} := -\boldsymbol{\varepsilon}^{t\,(k)}_{n+1} + \boldsymbol{\varepsilon}^t_n + \boldsymbol{\Lambda}^{(k)}_{n+1}[\xi^{(k)}_{n+1} - \xi_n] \tag{48}$$

and then linearizing them using Newton–Raphson iteration method [46, 47], where $\Phi_{n+1}^{(k)}$ and $\mathbf{R}_{n+1}^{t(k)}$ are ought to go to zero at the end of the iteration process. Equations (47) and (48) are linearized in the following way:

$$\Phi_{n+1}^{(k)} + \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \Delta\boldsymbol{\sigma}_{n+1}^{(k)} + \partial_{\xi}\Phi_{n+1}^{(k)}\Delta\xi_{n+1}^{(k)} = 0 \tag{49}$$

$$\mathbf{R}_{n+1}^{t(k)} - \Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)} + \boldsymbol{\Lambda}_{n+1}^{(k)}\Delta\xi_{n+1}^{(k)} + \begin{cases} (\xi_{n+1}^{(k)} - \xi_n)\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}_{n+1}^{(k)} : \Delta\boldsymbol{\sigma}_{n+1}^{(k)}, & \dot{\xi} > 0 \\ \mathbf{0}, & \dot{\xi} < 0 \end{cases} = 0 \tag{50}$$

where equation (50) is obtained by the Kuhn–Tucker condition when $(\xi_{n+1} - \xi_n) \neq 0$. Note that $\Delta\boldsymbol{\Lambda}_{n+1}^{(k)} = 0$ for reverse transformation because during reverse transformation $\boldsymbol{\Lambda}_{n+1}^{(k)}$ remains constant, for $k = 1, 2, \ldots$, with its components determined at the end of the previous forward transformation (cf. equation (7)). Equations (50) and (53) provide seven equations to solve for 13 unknowns, i.e. $\left\{\Delta\boldsymbol{\sigma}_{n+1}^{(k)}, \Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)}, \Delta\xi_{n+1}^{(k)}\right\}$.

The rest of the six equations are obtained by taking the increment of the stress-elastic strain state relation, equation (41), and expressing it for the total increment of strain as

$$\Delta\boldsymbol{\varepsilon}_{n+1} = \Delta\mathcal{S}_{n+1} : \boldsymbol{\sigma}_{n+1} + \mathcal{S}_{n+1} : \Delta\boldsymbol{\sigma}_{n+1} + \Delta\boldsymbol{\alpha}_{n+1}(T_{n+1} - T_0) + \boldsymbol{\alpha}_{n+1}\Delta T_{n+1} + \Delta\boldsymbol{\varepsilon}_{n+1}^{t} \tag{51}$$

Recalling the definitions of $\mathcal{S}$ and $\boldsymbol{\alpha}$ from equations (2a), (2b) and incrementing each, we get

$$\begin{aligned} \Delta\mathcal{S}_{n+1} &= \Delta\mathcal{S}\,\mathrm{d}\xi_{n+1} \\ d\boldsymbol{\alpha}_{n+1} &= \Delta\boldsymbol{\alpha}\,\mathrm{d}\xi_{n+1} \end{aligned} \tag{52}$$

where $\Delta\mathcal{S}$ and $\Delta\boldsymbol{\alpha}$ are defined in equation (12). Using equations (52) and noting that $\boldsymbol{\varepsilon}$ and T are fixed during the return mapping (transformation corrector) stage, equation (51) can be written in incremental form for $\Delta\boldsymbol{\varepsilon}_{n+1}^{t}$ at the kth iteration as follows:

$$\Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)} = -\mathcal{S}_{n+1}^{(k)} : \Delta\boldsymbol{\sigma}_{n+1}^{(k)} - [\Delta\mathcal{S} : \boldsymbol{\sigma}_{n+1}^{(k)} + \Delta\boldsymbol{\alpha}(T_{n+1} - T_0)]\Delta\xi_{n+1}^{(k)} \tag{53}$$

Equations (49), (50) and (53) form a system of equations to solve for $\left\{\Delta\boldsymbol{\sigma}_{n+1}^{(k)}, \Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)}, \Delta\xi_{n+1}^{(k)}\right\}$ for both forward and reverse transformations.

For the forward transformation ($\dot{\xi} > 0$), equation (49) is substituted into equation (53) to obtain

$$\Delta\boldsymbol{\sigma}_{n+1}^{(k)} = \mathscr{E}_{n+1}^{(k)} : \left[-\mathbf{R}_{n+1}^{t(k)} - \Delta\xi_{n+1}^{(k)}\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}\right] \tag{54}$$

where

$$\mathscr{E}_{n+1}^{(k)} := \left[\mathcal{S}_{n+1}^{(k)} + (\xi_{n+1}^{(k)} - \xi_n)\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}_{n+1}^{(k)}\right]^{-1} \tag{55}$$

and

$$\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} = \Delta\mathcal{S} : \boldsymbol{\sigma}_{n+1}^{(k)} + \Delta\boldsymbol{\alpha}(T_{n+1} - T_0) + \boldsymbol{\Lambda}_{n+1}^{(k)} \tag{56}$$

as shown in equation (20a). Substituting equation (54) into equation (50), the increment of the martensitic volume fraction is obtained for the kth iteration as

$$\Delta\xi_{n+1}^{(k)} = \frac{\Phi_{n+1}^{(k)} - \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathscr{E}_{n+1}^{(k)} : \mathbf{R}_{n+1}^{t(k)}}{\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathscr{E}_{n+1}^{(k)} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} - \partial_{\xi}\Phi_{n+1}^{(k)}} \tag{57}$$

A similar analysis can be performed for the reverse transformation ($\dot{\xi} < 0$) to obtain

$$\Delta\boldsymbol{\sigma}_{n+1}^{(k)} = \mathcal{S}_{n+1}^{(k)^{-1}} : \left[-\mathbf{R}_{n+1}^{t(k)} + \Delta\xi_{n+1}^{(k)}\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}\right] \tag{58}$$

Following the strategy to solve for $\Delta\xi_{n+1}^{(k)}$ in the forward transformation, we get

$$\Delta\xi_{n+1}^{(k)} = \frac{\Phi_{n+1}^{(k)} - \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathcal{S}_{n+1}^{(k)^{-1}} : \mathbf{R}_{n+1}^{t(k)}}{-\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathcal{S}_{n+1}^{(k)^{-1}} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} - \partial_{\xi}\Phi_{n+1}^{(k)}} \tag{59}$$

Furthermore the term $\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}_{n+1}^{(k)}$, appearing in the expression for $\mathscr{E}_{n+1}^{(k)}$, can be derived as

$$\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}_{n+1}^{(k)} = \sqrt{\frac{3}{2}}\frac{H}{\|\boldsymbol{\sigma}'\|}\left[\mathbf{I} - \frac{1}{3}\mathbf{1}\otimes\mathbf{1} - \frac{\boldsymbol{\sigma}'}{\|\boldsymbol{\sigma}'\|}\otimes\frac{\boldsymbol{\sigma}'}{\|\boldsymbol{\sigma}'\|}\right] \tag{60}$$

Here $\mathbf{I}$ is the most general rank four symmetric identity tensor and in indicial notation is given by

$$\mathbf{I} = \tfrac{1}{2}[\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}]\mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l \tag{61}$$

where $\otimes$ denotes the tensor product and $\mathbf{1}$ is the rank two identity tensor given by

$$\mathbf{1} = \delta_{ij}\mathbf{e}_i \otimes \mathbf{e}_j \tag{62}$$

In equations (53), (54) and (57) for forward transformation and equations (53), (58) and (59) for reverse transformation, we have a complete set of equations for the 13 unknown variables $\Delta\xi_{n+1}^{(k)}$, $\Delta\boldsymbol{\sigma}_{n+1}^{(k)}$ and $\Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)}$. The next step is to update the transformation strain, $\boldsymbol{\varepsilon}_{n+1}^{t(k)}$, and the martensitic volume fraction, $\xi_{n+1}^{(k)}$, by

$$\begin{aligned} \boldsymbol{\varepsilon}_{n+1}^{t(k+1)} &= \boldsymbol{\varepsilon}_{n+1}^{t(k)} + \Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)} \\ \xi_{n+1}^{(k+1)} &= \xi_{n+1}^{(k)} + \Delta\xi_{n+1}^{(k)} \end{aligned} \tag{63}$$

It is important to notice here that this is an implicit algorithm involving solution of 6×6 system of equations. A systematic procedure for carrying out the algorithm is outlined in Box 1.

Box 1. Closest point projection implicit algorithm for shape memory alloy constitutive models

1. Let $k=0,\ \xi_{n+1}^{(0)}=\xi_n,\ \boldsymbol{\varepsilon}_{n+1}^{t(0)}=\boldsymbol{\varepsilon}_n^t,\ \mathcal{S}_{n+1}^{(0)}=\mathcal{S}_n,\ \boldsymbol{\alpha}_{n+1}^{(0)}=\boldsymbol{\alpha}_n$
2. Calculate the thermoelastic prediction, and evaluate the transformation condition and transformation strain residual

$$\boldsymbol{\sigma}_{n+1}^{(k)} := \mathcal{S}_{n+1}^{(k)^{-1}} : \left[\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_{n+1}^{(k)}(T_{n+1}-T_o) - \boldsymbol{\varepsilon}_{n+1}^{t(k)}\right]$$

$$\Phi_{n+1}^{(k)} := \Phi\left[\boldsymbol{\sigma}_{n+1}^{(k)}, T_{n+1}, \xi_{n+1}^{(k)}\right]$$

$$\mathbf{R}_{n+1}^{t(k)} := -\boldsymbol{\varepsilon}_{n+1}^{t(k)} + \boldsymbol{\varepsilon}_n^t + \boldsymbol{\Lambda}_{n+1}^{(k)}\left[\xi_{n+1}^{(k)} - \xi_n\right]$$

If $|\Phi_{n+1}^{(k)}| \leqslant TOLERANCE1$ and $\|\mathbf{R}_{n+1}^{t(k)}\| \leqslant TOLERANCE2$ then
EXIT.
Else
forward or reverse transformation:

3. Compute elastic/algorithmic tangent moduli
Forward transformation ($\dot{\xi} > 0$):

$$\mathscr{E}_{n+1}^{(k)} := \left[\mathcal{S}_{n+1}^{(k)} + (\xi_{n+1}^{(k)} - \xi_n)\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}_{n+1}^{(k)}\right]^{-1}$$

Reverse transformation ($\dot{\xi} < 0$):
Invert $\mathcal{S}_{n+1}^{(k)}$
4. Compute increment of martensitic volume fraction, stress and transformation strain
Forward transformation ($\dot{\xi} > 0$):

$$\Delta\xi_{n+1}^{(k)} = \frac{\Phi_{n+1}^{(k)} - \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathscr{E}_{n+1}^{(k)} : \mathbf{R}_{n+1}^{t(k)}}{\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathscr{E}_{n+1}^{(k)} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} - \partial_{\xi}\Phi_{n+1}^{(k)}}$$

$$\Delta\boldsymbol{\sigma}_{n+1}^{(k)} = \mathscr{E}_{n+1}^{(k)} : \left[-\mathbf{R}_{n+1}^{t(k)} - \Delta\xi_{n+1}^{(k)}\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}\right]$$

Reverse transformation ($\dot{\xi} < 0$):

$$\Delta\xi_{n+1}^{(k)} = \frac{\Phi_{n+1}^{(k)} - \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathcal{S}_{n+1}^{(k)^{-1}} : \mathbf{R}_{n+1}^{t(k)}}{-\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathcal{S}_{n+1}^{(k)^{-1}} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} - \partial_{\xi}\Phi_{n+1}^{(k)}}$$

$$\Delta\boldsymbol{\sigma}_{n+1}^{(k)} = \mathcal{S}_{n+1}^{(k)^{-1}} : \left[-\mathbf{R}_{n+1}^{t(k)} + \Delta\xi_{n+1}^{(k)}\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}\right]$$

For either transformation:

$$\Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)} = -\mathcal{S}_{n+1}^{(k)} : \Delta\boldsymbol{\sigma}_{n+1}^{(k)} - \left[\Delta\mathcal{S} : \boldsymbol{\sigma}_{n+1}^{(k)} + \Delta\boldsymbol{\alpha}(T_{n+1}-T_0)\right]\Delta\xi_{n+1}^{(k)}$$

5. Update martensitic volume fraction and transformation strain

$$\xi_{n+1}^{(k+1)} = \xi_{n+1}^{(k)} + \Delta\xi_{n+1}^{(k)}$$

$$\boldsymbol{\varepsilon}_{n+1}^{t(k+1)} = \boldsymbol{\varepsilon}_{n+1}^{t(k)} + \Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)}$$

6. Update elastic and thermal moduli: $\mathcal{S}_{n+1}^{(k)^{-1}},\ \boldsymbol{\alpha}_{n+1}^{(k)}$
Let $k=k+1$ & go to 2

3.2.3. Geometrical interpretation. Before proceeding to derive the consistent tangent moduli tensors for the SMA constitutive model, let us understand the geometric meaning of the transformation correction in stress space. For the $(k+1)$th iteration at t_{n+1}, substituting equation (42) into equation (41), we get

$$\boldsymbol{\sigma}_{n+1}^{(k+1)} = \mathcal{S}_{n+1}^{(k+1)^{-1}} : \{\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_{n+1}^{(k+1)}(T_{n+1} - T_0) - \boldsymbol{\varepsilon}_n^t - [\xi_{n+1}^{(k+1)} - \xi_n]\boldsymbol{\Lambda}_{n+1}^{(k+1)}\} \tag{64}$$

Premultiplying equations (64) and (65) by $\mathcal{S}_{n+1}^{(k+1)}$ and $\mathcal{S}_n^{-1}$, respectively, and then subtracting the resulting equations, the expression for the stress at the end of the $(k+1)$th iteration in terms of the trial thermoelastic stress at t_{n+1} can be obtained after some algebraic manipulations as

$$\boldsymbol{\sigma}_{n+1}^{(k+1)} = \boldsymbol{\sigma}_{n+1}^{(0)} - \Delta\xi_{n+1}^{(k)}\mathcal{S}_{n+1}^{(k+1)^{-1}} : \{\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k+1)} - \Delta\mathcal{S} : [\boldsymbol{\sigma}_{n+1}^{(k)} - \boldsymbol{\sigma}_{n+1}^{(0)}]\} \tag{65}$$

The above equation describes the return (relaxation) path of the stress tensor from the thermoelastic prediction with each iteration. It is clear that the return path is not arbitrary but is defined by the projection of the thermoelastic prediction on the transformation surface, Φ_{n+1}^{k+1}, in a way defined by equation (65). That is, the stress state is updated during the iterative procedure in such a manner that at $(k+1)$th iteration, in the six-dimensional stress space defined by the metric of elasticities, $\mathcal{S}_{n+1}^{(k+1)^{-1}}$, $\boldsymbol{\sigma}_{n+1}^{(k+1)}$ lies on Φ_{n+1}^{k+1}, and the vector joining $\boldsymbol{\sigma}_{n+1}^{(k+1)}$ and $\boldsymbol{\sigma}_{n+1}^{(0)}$ is normal to Φ_{n+1}^{k+1} for the case when $\Delta\mathcal{S} = 0$. That is, if the change in material properties is not significant, the stress will be updated normal to each iterative transformation surface. Note that the implicit integration of the transformation strain flow rule results in the above-mentioned characteristic of the return path.

3.2.4. Consistent tangent moduli tensors. As already shown, the solution to the rate type SMA constitutive equations require numerical integration over a number of time steps. The stress tensor, as a result of this process, is defined as a function of the deformation history up to a given instant. The crucial next step is that the tangent tensors that appear in the linearized problem (due to Newton's method) must be obtained by employing the response function associated with the integration algorithm. This is important if quadratic rate of asymptotic convergence is to be maintained. In other words, the notion of consistent tangent stiffness tensor and tangent thermal moduli tensor arises due to the enforcement of the consistency condition on the discrete algorithmic problem. On the other hand, the continuum tangent stiffness and tangent thermal moduli tensors result from the enforcement of the classical consistency condition on the continuum problem. Simo and Taylor [43] showed for the case of associative/non-associative elasto-plasticity using numerical examples that the use of the continuum tangent stiffness leads to loss of the quadratic rate of asymptotic convergence which characterizes Newton's iteration method. In this section, only the final result for forward transformation is presented and the detailed derivation is given in the appendix. The consistent tangent tensors for the reverse transformation are exactly the same as the continuum tangent tensors given in equations (32) and (33).

For the forward phase transformation, the algorithmic tangent stiffness tensor, $\mathscr{L}$, and the algorithmic tangent thermal moduli tensor, $\boldsymbol{\Theta}$, are derived as

$$\mathscr{L} = \mathscr{E}_{n+1} - \mathbf{H}_{n+1} \otimes \mathbf{H}_{n+1} \tag{66}$$

$$\boldsymbol{\Theta} = -\mathscr{L} : \boldsymbol{\alpha}_{n+1} - \frac{\partial_T \Phi_{n+1}}{c}\mathbf{H}_{n+1} \tag{67}$$

where

$$
\begin{aligned}
c &:= \sqrt{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1} - \partial_{\xi}\Phi_{n+1}} \\
\mathbf{H}_{n+1} &:= \frac{\mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}}{\sqrt{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1} - \partial_{\xi}\Phi_{n+1}}} = \frac{\mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}}{c} \\
\mathscr{E}_{n+1} &:= [\mathcal{S}_{n+1} + (\xi_{n+1} - \xi_n)\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}_{n+1}]^{-1}
\end{aligned}
\tag{68}
$$

Now if expressions for the continuum and consistent tangent tensors are compared for forward transformation, it is observed that the difference between the tangent tensors (equations (28), (29) and (66), (67)) arises due to the algorithmic stiffness tensor defined in equation (68). It is observed that as the time step $\Delta t = t_{n+1} - t_n \to 0$, $\xi_{n+1} - \xi_n \to 0^+$, i.e. as

$$
\Delta t \to 0 \Rightarrow \begin{cases} \mathscr{E} \to \mathcal{S}^{-1} \\ \mathbf{H} \to \mathbf{A} \\ c \to a \end{cases}
\tag{69}
$$

where $(\mathbf{A}, a)$ and $(\mathbf{H}, c)$ are defined in equations (30) and (68), respectively. Therefore, the consistent tangent stiffness tensor and consistent tangent thermal moduli tensor reduce to their continuum counterparts as $\Delta t \to 0$ for the forward transformation. This shows that the closest point projection algorithm is consistent with the continuum problem. It is also observed that for large time (loading) increments, much transformation may result in ξ_{n+1} being much larger than ξ_n. This will result in significant deviation of the algorithmic tensors from the continuum tensors.

For the reverse transformation, the consistent tangent stiffness tensor and consistent tangent thermal moduli tensor are the same as their continuum counterparts.

3.3. Cutting plane return mapping algorithm

As it is noticed in the closest point algorithm, the usage of backward Euler integration leads to systems of non-linear equations, whose solution by Newton's iteration method requires evaluation of the gradients of transformation flow direction. Addition of another internal state variable in the model (e.g. [2]) will require further gradients to be accounted for. The task of evaluating such quantities may prove difficult and computationally expensive. The cutting plane return mapping algorithm is proposed to avoid the need for computing the above-mentioned gradients [30, 42]. The present algorithm is formulated solely on the basis of the transformation function, the consistency condition, direction of the transformation flow and the elastic stiffness tensors without evaluating any gradient. The analysis falls within the class of convex cutting plane methods of constrained optimization and utilizes the operator splitting methodology [46].

The basic idea relies on integrating the transformation correction equations (equation (39)) in an explicit manner and linearizing the consistency condition. The transformation residual, $\mathbf{R}^{t(k)}_{n+1}$, and the terms containing the derivatives of the transformation tensor, $\boldsymbol{\Lambda}^{(k)}_{n+1}$, do not appear in the formulation. The Newton's iteration method is only applied to calculate the increment of martensitic volume fraction, $\Delta\xi^{(k)}_{n+1}$. This enables the convergence of the algorithm towards the final value of the state variables at a quadratic rate; but, convergence is not guaranteed. The thermoelastic part of the algorithm is the same as given in Section 3.2.1; therefore, the focus will be on the transformation correction which is the main difference between the two algorithms.

3.3.1. Transformation correction. Here we assume transformation loading; that is, the thermoelastic prediction does not satisfy the transformation condition and $\Phi_{n+1}^{(0)} > 0$. Recalling that the total strain tensor, $\boldsymbol{\varepsilon}$, and temperature, T, are fixed, the transformation flow rule given by equation (7) is discretized explicitly as follows:

$$\Delta\boldsymbol{\varepsilon}_{n+1}^{t\,(k)} = \Delta\xi_{n+1}^{(k)}\boldsymbol{\Lambda}_{n+1}^{(k)} \tag{70}$$

Note here that the above equation can also be obtained from equation (50) in Section 3.2 by neglecting the residual, $\mathbf{R}_{n+1}^{t(k)}$ and the last term containing gradient of the transformation tensor, $\boldsymbol{\Lambda}_{n+1}^{(k)}$. Also, this explicit integration is being carried out from the initial thermoelastic state to the final unknown state. Now using the discretized flow rule, equations (53) and (56), the stress increment for the kth iteration during the $(n+1)$th increment can be written as

$$\Delta\boldsymbol{\sigma}_{n+1}^{(k)} = -\,\mathcal{S}_{n+1}^{(k)^{-1}}\Delta\xi_{n+1}^{(k)}\begin{cases}\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}, & \dot{\xi} > 0\\ -\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}, & \dot{\xi} < 0\end{cases} \tag{71}$$

Linearizing the discrete transformation function for the kth iteration and equating it to zero due to the Kuhn–Tucker condition, i.e. $\xi_{n+1} - \xi_n \neq 0$, we get

$$\Phi_{n+1}^{(k)} + \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \Delta\boldsymbol{\sigma}_{n+1}^{(k)} + \partial_{\xi}\Phi_{n+1}^{(k)}\Delta\xi_{n+1}^{(k)} = 0 \tag{72}$$

Substituting equation (71) into equation (72) and solving for the increment of ξ, the following expression is obtained:

$$\Delta\xi_{n+1}^{(k)} = \frac{\Phi_{n+1}^{(k)}}{\pm\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathcal{S}_{n+1}^{(k)^{-1}} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} - \partial_{\xi}\Phi_{n+1}^{(k)}} \tag{73}$$

where the $+$ sign is for the forward transformation and $-$ sign for the reverse transformation. In equations (73), (71) and (70), we have a complete set of equations in terms of the 13 variables $\Delta\xi_{n+1}^{(k)}$, $\Delta\boldsymbol{\sigma}_{n+1}^{(k)}$ and $\Delta\boldsymbol{\varepsilon}_{n+1}^{t\,(k)}$ for both forward and reverse transformations. Note that unlike the closest point formulation, the incremental expressions are the same for both the forward and reverse transformations, except the change of sign in equations (71) and (73), thus saving a lot of coding details. The next step is to update the transformation strain, $\boldsymbol{\varepsilon}_{n+1}^{t\,(k)}$, and the martensitic volume fraction, $\xi_{n+1}^{(k)}$, as given below:

$$\begin{aligned}\boldsymbol{\varepsilon}_{n+1}^{t(k+1)} &= \boldsymbol{\varepsilon}_{n+1}^{t\,(k)} + \Delta\boldsymbol{\varepsilon}_{n+1}^{t\,(k)}\\ \xi_{n+1}^{(k+1)} &= \xi_{n+1}^{(k)} + \Delta\xi_{n+1}^{(k)}\end{aligned} \tag{74}$$

For the cutting plane algorithm, the continuum tangent tensors given in Section 2.1 have to be used since consistent tangent tensors cannot be obtained in closed form. This means a loss in the quadratic convergence rate of Newton's iteration method, and therefore, quasi-Newton methods are suggested for global solution strategies [46]. A detailed strategy for implementing the algorithm is given in Box 2.

Box 2. Convex cutting plane explicit algorithm for shape memory alloy constitutive models

1. Let $k=0,\ \xi_{n+1}^{(0)}=\xi_n,\ \boldsymbol{\varepsilon}_{n+1}^{t\ (0)}=\boldsymbol{\varepsilon}_n^t,\ \mathcal{S}_{n+1}^{(0)}=\mathcal{S}_n,\ \boldsymbol{\alpha}_{n+1}^{(0)}=\boldsymbol{\alpha}_n$

2. Calculate the thermoelastic prediction and evaluate the transformation condition

$$\boldsymbol{\sigma}_{n+1}^{(k)} := \mathcal{S}_{n+1}^{(k)^{-1}} : [\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_{n+1}^{(k)}(T_{n+1}-T_o) - \boldsymbol{\varepsilon}_{n+1}^{t(k)}]$$
$$\Phi_{n+1}^{(k)} := \Phi[\boldsymbol{\sigma}_{n+1}^{(k)}, T_{n+1}, \xi_{n+1}^{(k)}]$$

   If $|\Phi_{n+1}^{(k)}| \leqslant$ *TOLERANCE* then

   EXIT.

   Else

   forward or reverse transformation:

3. Compute increment of martensitic volume fraction and transformation strain

$$\Delta\xi_{n+1}^{(k)} = \frac{\Phi_{n+1}^{(k)}}{\pm\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} : \mathcal{S}_{n+1}^{(k)^{-1}} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)} - \partial_{\xi}\Phi_{n+1}^{(k)}}$$
$$\Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)} = \Delta\xi_{n+1}^{(k)}\boldsymbol{\Lambda}_{n+1}^{(k)}$$

   where $\pm$ stands for forward and reverse transformations, respectively.

4. Update martensitic volume fraction and transformation strain

$$\xi_{n+1}^{(k+1)} = \xi_{n+1}^{(k)} + \Delta\xi_{n+1}^{(k)}$$
$$\boldsymbol{\varepsilon}_{n+1}^{t\ (k+1)} = \boldsymbol{\varepsilon}_{n+1}^{t(k)} + \Delta\boldsymbol{\varepsilon}_{n+1}^{t(k)}$$

   Let $k=k+1$ & go to 2

3.3.2. Geometrical interpretation. A geometric interpretation of the algorithm can be given by expanding equation (71) in the following form:

$$\boldsymbol{\sigma}_{n+1}^{(k+1)} = \boldsymbol{\sigma}_{n+1}^{(k)} - \Delta\xi_{n+1}^{(k)}\mathcal{S}_{n+1}^{(k)^{-1}} : \begin{cases} \partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}, & \dot{\xi}>0 \\ -\partial_{\boldsymbol{\sigma}}\Phi_{n+1}^{(k)}, & \dot{\xi}<0 \end{cases} \tag{75}$$

The above expression implies that during the transformation correction $\boldsymbol{\sigma}_{n+1}^{(k+1)}$ is found by projecting the plane normal to $\Phi_{n+1}^{(k)}$ and finding its intersection with level iterate $\Phi_{n+1}^{(k+1)}$. That is, the return path is approximated by a sequence of straight segments. Note that the explicit integration of the transformation strain flow rule is directly responsible for the above-mentioned characteristics of the return path.

3.4. Summary of algorithms

Having derived the algorithms, attention is now turned to their respective features. Following are the remarks on the closest point projection algorithm:

1. It is based on the backward Euler integration rule of the transformation strain flow rule resulting in a set of non-linear algebraic equations solved using Newton's iteration method.
2. The quadratic convergence rate from the thermoelastic prediction to the final solution of the Newton's iteration method is guaranteed, as observed in the variation of the transformation function with changing martensitic volume fraction during transformation correction in Figures 4 and 5.
3. Backward Euler integrations makes the local integration procedure first-order accurate and unconditionally stable.
4. The algorithm facilitates the derivation of the consistent tangent moduli tensors in closed form, thereby preserving the global quadratic convergence rate.
5. The algorithm requires the solution of a system of linear algebraic relations (equations (49), (50) and (53)). This results in a relatively large number of tensorial operations beside the evaluation of the gradient of the transformation tensor (equation (60)) and inversion of the 6×6 algorithmic tangent tensor (equation (55)) during forward transformation.

Similar remarks on the features of the convex cutting plane algorithm are as follows:

1. It is based on the explicit integration of the transformation flow rule from the thermoelastic predictor; the satisfaction of the transformation condition still requires the use of Newton's iteration method.
2. The convergence of the algorithm towards the final value of the state variables is obtained at a quadratic rate as shown in Figures 4 and 5.
3. Forward Euler integration makes the load-stepping procedure first-order accurate. The algorithm is also unconditionally stable in the forward transformation because the iteration function to find the increment of martensitic volume fraction, $\Delta\xi$, is contractive and the transformation tensor is normalized (cf. Section 2). In the reverse transformation, Newton's method is only locally convergent which may result in instability.
4. The algorithm does not allow the derivation of consistent tangent moduli tensors in closed form, thus requiring continuum tangent moduli tensors to be used. This may require usage of quasi-Newton techniques for faster convergence in the global solution scheme.
5. Its fundamental advantage is its simplicity, since it does not require computation of the gradient of the transformation tensor or inversion of algorithmic tangent tensors, and only involves a few tensorial operations and function evaluations to determine the unknowns.

4. RESULTS AND DISCUSSIONS

To characterize the behaviour of the two algorithms, five boundary value problems (BVPs) are analysed. Three simpler uniaxial cases are studied to characterize the evolution of the transformation and stress state in a given load increment. Only two 3-D eight node linear elements are used to model each of these problems. The other two torsional and combined loading cases are performed to analyse the numerical accuracy and computation time. Special eight-node quadratic axisymmetric elements with angle of twist as a degree of freedom are used to discretize these problem domains.

Table I. Shape memory alloy material parameters

Material constants	Values	Model variables
E^{A}	$70{\cdot}0 \times 10^9$ Pa	Used to calculate isotropic compliance tensors, $\mathcal{S}^{\mathrm{A}}$ and $\mathcal{S}^{\mathrm{M}}$
E^{M}	$30{\cdot}0 \times 10^9$ Pa	
$\nu^{\mathrm{A}} = \nu^{\mathrm{M}}$	0·3	
α^{A}	$22{\cdot}0 \times 10^{-6}$/K	Used to calculate isotropic thermal expansion coefficient tensors, $\boldsymbol{\alpha}^{\mathrm{A}}$ and $\boldsymbol{\alpha}^{\mathrm{M}}$
α^{M}	$10{\cdot}0 \times 10^{-6}$/K	
$\rho\Delta c = c^{\mathrm{M}} - c^{\mathrm{A}}$	0·0 J/(m^3 K)	
H	0·05	$\dfrac{\rho\Delta s_0}{H} = -\left(\dfrac{\mathrm{d}\sigma}{\mathrm{d}T}\right)^{\mathrm{A}}$
$\left(\dfrac{\mathrm{d}\sigma}{\mathrm{d}T}\right)^{\mathrm{A}} = \left(\dfrac{\mathrm{d}\sigma}{\mathrm{d}T}\right)^{\mathrm{M}}$	$7{\cdot}0 \times 10^6$ Pa/(m^{3}K)	
$A^{\circ\mathrm{f}}$	315·0 K	$\gamma = \rho\Delta u_0 + \mu_1 = \frac{1}{2}\rho\Delta s_0(M^{\circ\mathrm{s}} + A^{\circ\mathrm{f}})$
$A^{\circ\mathrm{s}}$	295·0 K	$\rho b^{\mathrm{A}} = -\rho\Delta s_0(A^{\circ\mathrm{f}} - A^{\circ\mathrm{s}})$
$M^{\circ\mathrm{s}}$	291·0 K	$\rho b^{\mathrm{M}} = -\rho\Delta s_0(M^{\circ\mathrm{s}} - M^{\circ\mathrm{f}})$
$M^{\circ\mathrm{f}}$	271·0 K	$\mu_2 = \frac{1}{4}(\rho b^{\mathrm{A}} - \rho b^{\mathrm{M}})$
		$Y^* = -\frac{1}{2}\rho\Delta s_0(A^{\circ\mathrm{f}} - M^{\circ\mathrm{s}}) + \frac{1}{4}\rho\Delta s_0(M^{\circ\mathrm{s}} - M^{\circ\mathrm{f}} - A^{\circ\mathrm{f}} + A^{\circ\mathrm{s}})$

A = austenite, M = martensite

The boundary value problems are numerically simulated using the commercially available non-linear finite element code ABAQUS on a parallel processing SGI Power Challenge composed of 24 RL10000 processors. The analyses are run using the 'small' queue option which allows usage of no more than two processors simultaneously. The incremental (discretized) SMA constitutive model based on the two algorithms is implemented in the user supplied subroutine UMAT. The material properties used in the FE solution of the boundary value problems are given in Table I.

In order to be able to compare the results from the implementation of both algorithms, only the transformation function criterion value (cf. point 2 of Boxes 1 and 2) is selected to end the iteration procedure in the transformation correction part. Therefore, the same value of tolerance $1{\cdot}0 \times 10^{-9}$ MPa is chosen for both algorithms. This leads to six significant digits for $\boldsymbol{\varepsilon}^t$ and nine significant digits for ξ. All the results discussed herein are obtained at the mid-point of a finite element, a task performed by ABAQUS and they are feasible because most of the following problems result in uniform stress state throughout the domain. The results and discussion of these problems follow.

4.1. Uniaxial cases

4.1.1. Uniaxial isothermal pseudoelasticity. Figure 6 shows the schematic of the BVP. The initial temperature is chosen to be $A^{\circ\mathrm{f}}$ (see Table I). The boundary conditions are such that except for the axial component of stress σ_{11} the rest of the stresses $\sigma_{ij} = 0,\ i = 1,\ldots,3,\ j = 1,\ldots,3,\ i = j \neq 1)$ are zero. The loading history as shown in Figure 6 is applied equally at the four nodes to obtain constant axial stress throughout the domain and assumed to be a linear function of the loading parameter, t, where $t \in [0,2]$. The maximum load, $F_{\max}$, is applied such that all austenite transforms into martensite producing a transformation strain of 5 per cent followed by full unloading

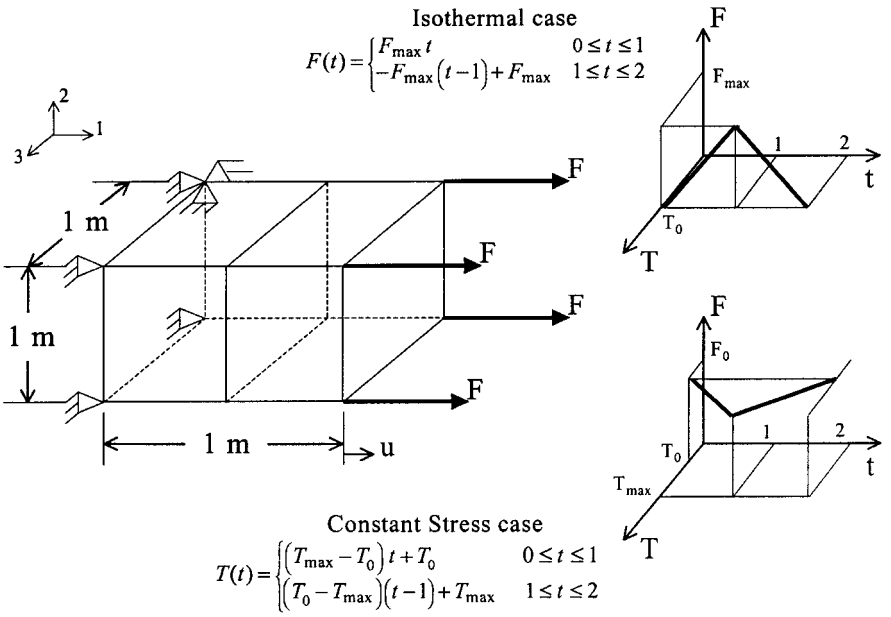


Figure 6. Schematic and loading histories of the first and second boundary value problem. The mechanical boundary conditions result in uniaxial stress in the 1-direction

to austenitic state (pseudoelastic effect). In Figure 7, the axial stress is plotted against the axial strain for both the convex cutting plane and the closest point projection algorithms. It is observed that they both match satisfactorily in describing the pseudoelasticity effect. In the following, until otherwise mentioned, all results presented will be equally valid for both algorithms at least up to the fourth decimal for all state variables.

To demonstrate how the transformation surface evolves during a typical return mapping (transformation correction) part of the algorithm, arbitrary load increments are chosen for the forward and reverse transformations, respectively. Figures 8 and 9 show the evolution of the transformation surface with the number of iterations within the $n+1$ increment, in the $\sigma_{11}-\sigma_{22}$ space for forward and reverse transformations, respectively. The magnitude of stresses with respect to iterations during the return mapping is also tabulated and given in the figures. The converged values of the martensitic volume fraction are $\xi_{n+1}=0{\cdot}7974$ and $\xi_{n+1}=0{\cdot}2776$ for forward and reverse transformations, respectively. The total number of iterations used to satisfy the prescribed tolerance are usually less than 6 for both forward and reverse transformations. As observed in the figures, even though the applied load is expected to result in an uniaxial stress state, this is not the case in the initial iterations of the transformation correction. This is due to the thermoelastic prediction which assumes initially that the transverse strains are totally elastic. This is corrected during the iterations by the algorithms and the final stress state obtained is found to be uniaxial. For the case of forward transformation as shown in Figure 8, the transformation function expands with number of iterations due to the increase in martensitic volume fraction, whereas the stress state evolves towards the transformation surface from the convex side. On the other hand, in the case of reverse transformation, the thermoelastic prediction lies on the concave side of the transformation surface

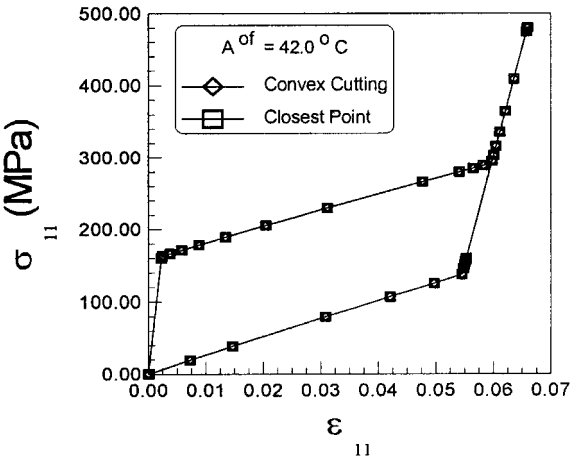


Figure 7. Axial stress versus axial strain hysteresis curve for full forward and then reverse transformation under uniformly applied load—Pseudoelasticity (Problem 1)

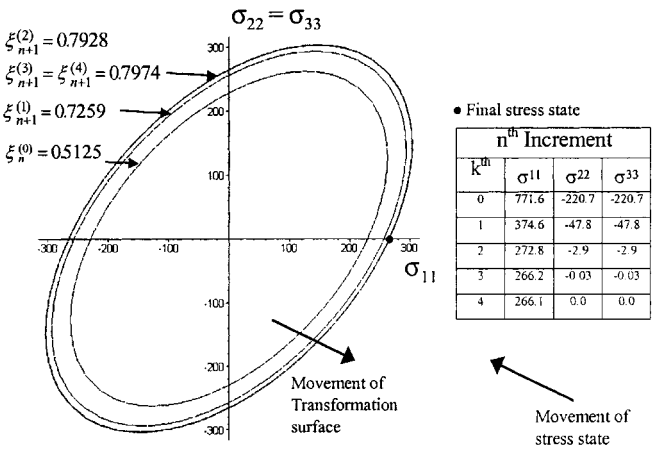


Figure 8. Evolution of forward transformation surface in $\sigma_{11}-\sigma_{22}$ space with return mapping iterations (k) for a given strain increment (problem 1). Note that the stresses do not lie on the transformation surface until the last iteration

in Figure 9. In subsequent iterations the stress state moves into the convex side of the transformation surface, and both the stress states and the transformation surface move towards each other until convergence is achieved.

The behaviour of the transformation surface discussed above is explained by observing Figures 10(a) and 10(b), where the transformation function is plotted against the martensitic volume fraction for both forward- and reverse-phase transformations for both algorithms. Figure 10(a) is the classical example of the application of Newton's method to finding the zero of a convex function, in this case the transformation function. This is the reason that Newton's method is globally convergent for return mapping procedure during forward transformation. This figure also

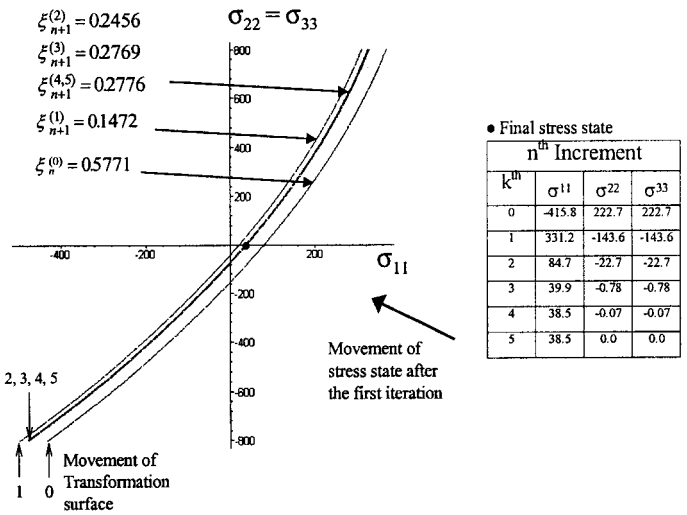


Figure 9. Evolution of reverse transformation surface in σ_{11}–σ_{22} space with return mapping iterations (k) for a given strain increment (problem 1). Note that the stresses do not lie on the transformation surface until the last iteration

shows that with each iteration, the martensitic volume fraction increases thus expanding the transformation surface in the stress space. The solid line in the graph is a fit drawn for the sake of clarity. In reverse transformation, however, as it is observed in Figure 10(b), the first Newton iteration results in a switch of sign for the transformation function value and then it converges just like the forward transformation. This also results in one more iteration to converge than those required for the forward transformation.

4.1.2. Uniaxial constant stress thermally induced transformation. In this BVP the SMA material is put through a temperature cycle at a constant axial stress of 50 MPa. The schematic shown in Figure 6 exhibit the boundary conditions along with the loading history. The stressed material is initially at a temperature above A^{of}. Next the temperature is decreased below M^{of} for full forward transformation and then increased above A^{of} again for full reverse transformation. In Figure 11, the uniaxial strain is plotted against applied temperature. The two algorithms again match closely. A transformation strain of 5 per cent is obtained at the end of the cooling part of the cycle, which fully recovers at the end of heating.

The evolution of transformation surface for this case is analysed in both the σ_{11}–T and σ_{11}–σ_{22} spaces for both forward and reverse transformation in Figures 12(a), 13, 12(b) and 14, respectively. In the forward transformation as seen in the σ_{11}–T space (Figure 12a), the transformation curve moves toward the imposed (lower) temperature of 6·5°C with the increase in the martensitic volume fraction with the number of iteration. On the other hand, the axial stress decreases from the thermoelastic overprediction until in finally converges to 50 MPa. Whereas in the σ_{11}–σ_{22} space (Figure 13), the transformation function expands with the number of iterations due to the increase in martensitic volume fraction, and the stress state evolves towards the transformation surface from the convex side until they finally converge at 50 MPa and 6·5°C. The evolution of stresses is also tabulated in Figure 13 which shows the rapid convergence. At the end of the transformation correction, ξ changes from 0·5930 to 0·9395.

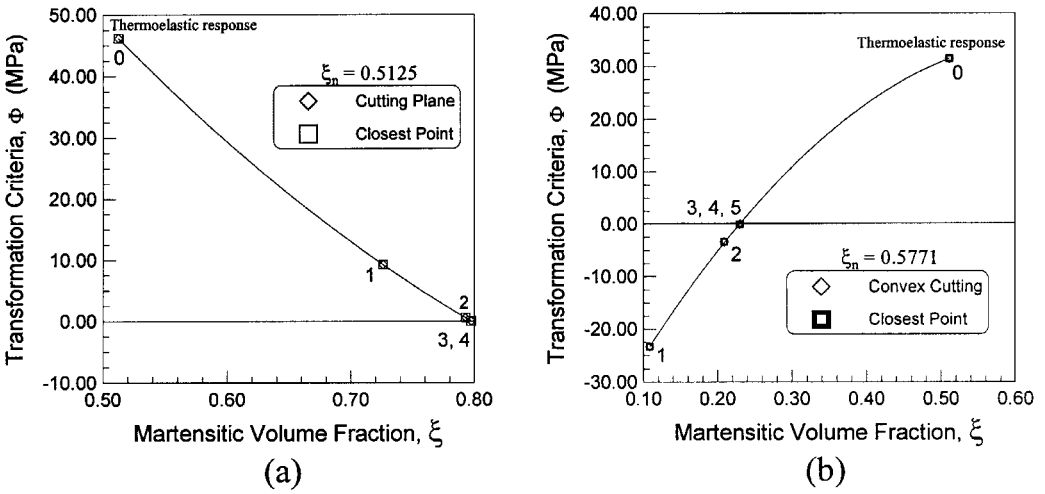


Figure 10. (a) Convergence of Φ to zero with increasing ξ in forward transformation correction; (b) convergence of Φ to zero with changing ξ in reverse transformation correction. Note the switch in the sign of Φ after the first iteration (problem 1)

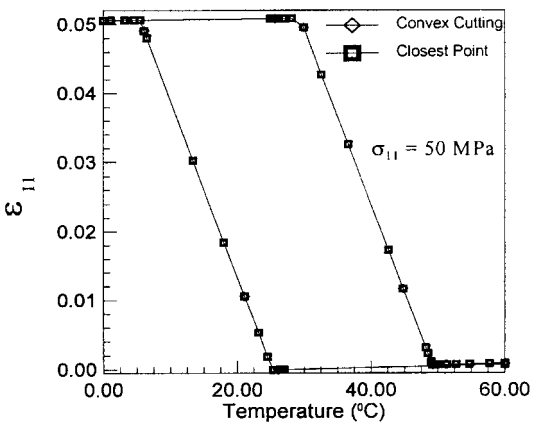


Figure 11. Axial strain versus temperature hysteresis curve (Problem 2)

Figure 12(b) contains the reverse transformation correction in σ_{11}–T space. For the first iteration the transformation curve overshoots the imposed (higher) temperature of 50·3°C and then in subsequent iterations it converts back. The axial stress shows a similar behaviour where the thermoelastic prediction results in compressive axial stress but changes sign in the first iteration, and similar to forward transformation, it decreases to 50 MPa at the end of the iterations. This can be observed in Figure 14 where the thermoelastic prediction switches from the concave side of the transformation surface into the convex side in the first iteration. In subsequent iterations, the stress state and the transformation surface move toward each other until convergence is achieved at 50 MPa and 50·3°C with ξ changing from 0·6320 to 0·3325. The evolution of stresses is also tabulated in Figure 14.

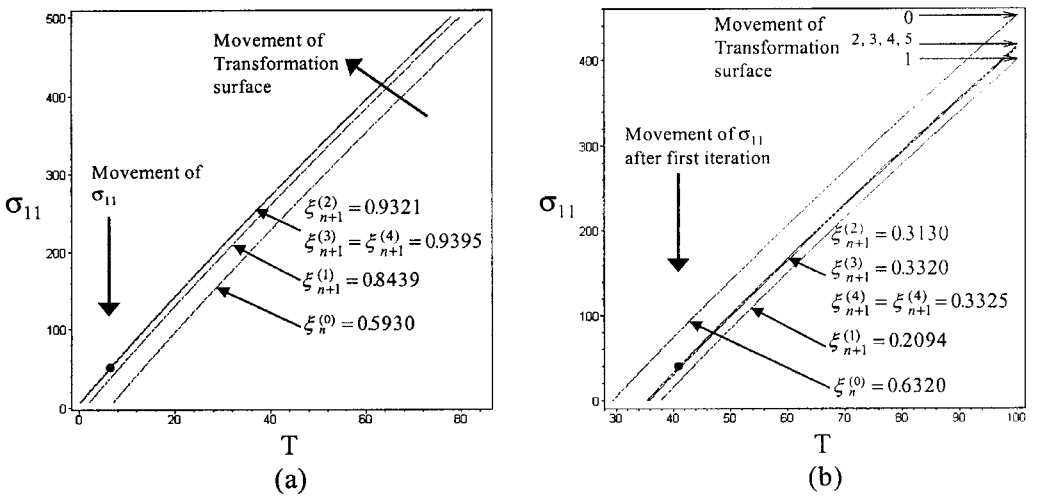
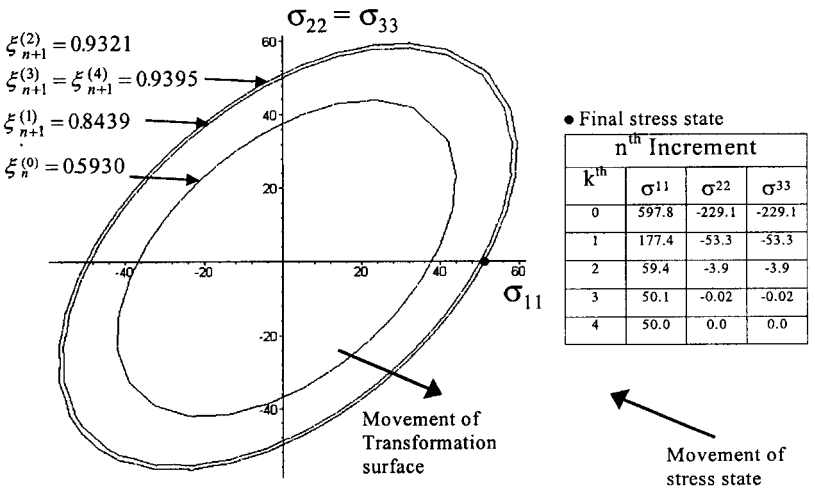


Figure 12. Evolution of (a) forward transformation surface; (b) reverse transformation surface in σ_{11}–T space with return mapping iterations (k) for given temperature and strain increments (problem 2)



n^{th} Increment			
k^{th}	σ^{11}	σ^{22}	σ^{33}
0	597.8	-229.1	-229.1
1	177.4	-53.3	-53.3
2	59.4	-3.9	-3.9
3	50.1	-0.02	-0.02
4	50.0	0.0	0.0

Figure 13. Evolution of forward transformation surface in σ_{11}–σ_{22} space with return mapping iterations (k) for a given strain increment (problem 2)

Once again, the above results can be described more clearly by plotting the transformation function against the martensitic volume fraction in Figures 15(a) and 15(b). As in the case of stress-induced pseudoelasticity, ξ increases for each forward transformation iteration; however, it decreases for the first iteration and then increases until the last iteration in the case of reverse transformation. Both times, convergence is easily achieved within five iterations and the reverse transformation takes one more iteration to converge due to the change in sign of the transformation function.

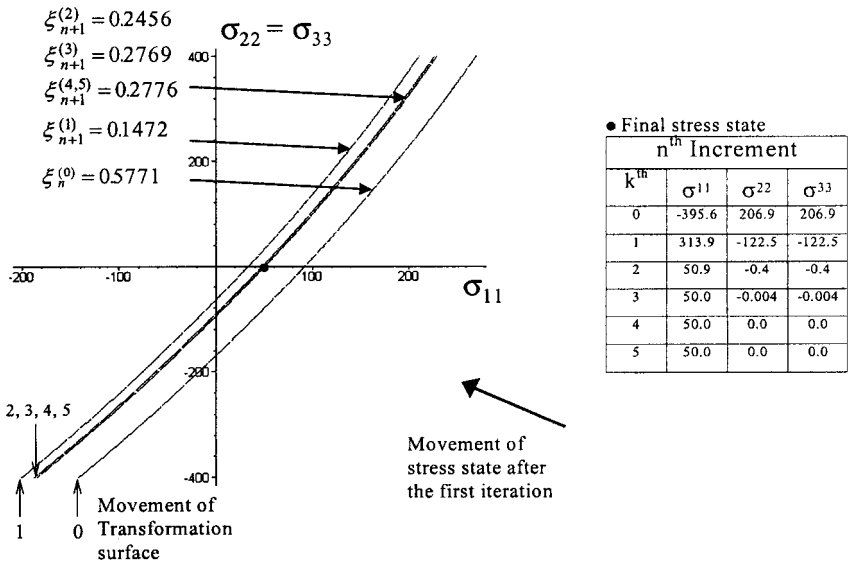


Figure 14. Evolution of reverse transformation surface in $\sigma_{11}-\sigma_{22}$ space with return mapping iterations (k) for a given strain increment (problem 2)

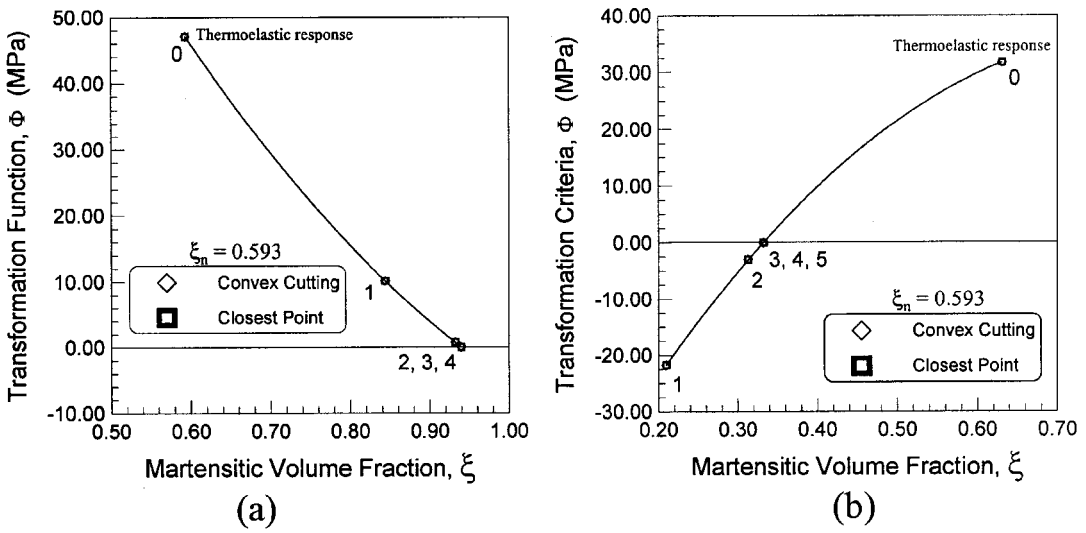


Figure 15. (a) Φ versus ξ during forward transformation; (b) Φ versus ξ during reverse transformation (problem 2)

4.1.3. Uniaxial thermally induced transformation with a spring. The BVP is similar to the first two in the sense of geometry and displacement boundary conditions. The resultant stress state is uniaxial (only σ_{11}). The differences arises from the placement of linear springs of equal stiffness ($k = 2.0 \times 10^9$ N/m) at the end of the SMA instead of point forces, as shown in Figure 16. The loading history is also given in the same figure. Initially, the material is in a stress free state at a

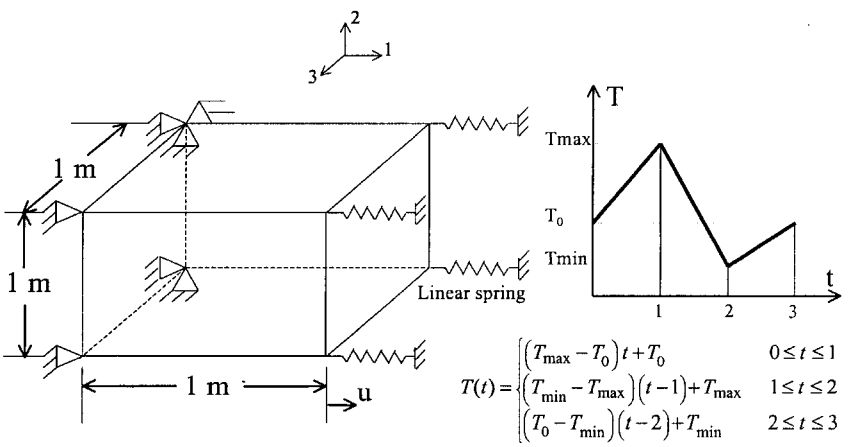


Figure 16. Schematic and loading history of the third boundary value problem

temperature below $A^{\circ s}$ and then the temperature is increased until the whole material transforms completely from martensite to austenite. At this point, the temperature is first decreased below $M^{\circ f}$ resulting in complete transformation from austenite to martensite and then increased to the initial value to complete the cycle. In Figure 17(a). σ_{11} is plotted as a function of temperature for both algorithms, which give similar results. This figure shows that as the temperature is increased, the transformation strain is recovered under a build-up of tensile stress due to the resistance of the springs. This results in the shifting of the austenitic finish temperature to higher values. Upon cooling the SMA, forward phase transformation results in transformation strain and thus relaxes the applied tensile stress on the SMA by the springs. At the end of the cycle, the material returns to its initial stress free state. In Figure 17(b), σ_{11} is plotted versus ε_{11} and the result from both algorithms coincide. Note that the linearity in the stress–strain response is due to the linear behaviour of the springs.

In Figures 18(a) and 18(b), the transformation function, $\Phi(\boldsymbol{\sigma}, \xi)$, is plotted versus the martensitic volume fraction, ξ, for three SMA constitutive models for both forward and reverse transformations in order to study their convergence behaviour during the transformation correction. All these models can be described under the framework of the unified SMA constitutive model. To plot the transformation function, the transformation flow rule is discretized using the forward Euler integration rule. The models chosen are the polynomial model [20], exponential model [17, 18], and cosine model [19]. It is noted in Figure 18(a) that $\Phi(\boldsymbol{\sigma}, \xi)$ decreases monotonically with increasing ξ in the forward transformation with a negative slope, as derived previously for the polynomial model in Section 3.1. This result, plus the fact that the transformation function is convex for forward transformation, guarantees global convergence for a solution by Newton's iteration method for all models. Similarly, in Figure 18(b), a plot of $\Phi(\boldsymbol{\sigma}, \xi)$ vs. ξ is shown for reverse transformation, where the transformation function is monotonically reducing with decreasing ξ with a positive slope, again as derived previously for the polynomial model in Section 3.1. This feature alone guarantees only local convergence of a Newton's iteration method because the transformation function is not convex. For a suitable load increment, as shown in Figure 5 after the first iteration, the transformation function does become convex and the method converges thereafter.

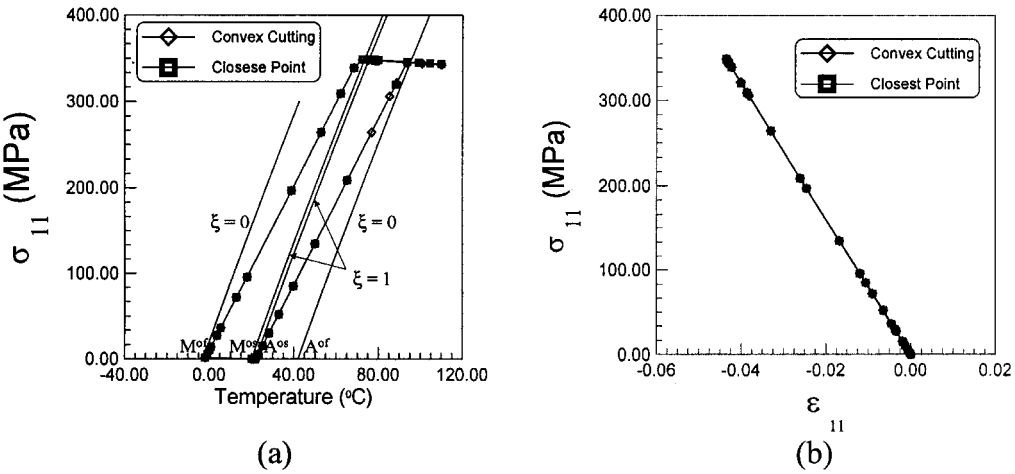


Figure 17. Axial stress versus (a) temperature; (b) axial strain (problem 3)

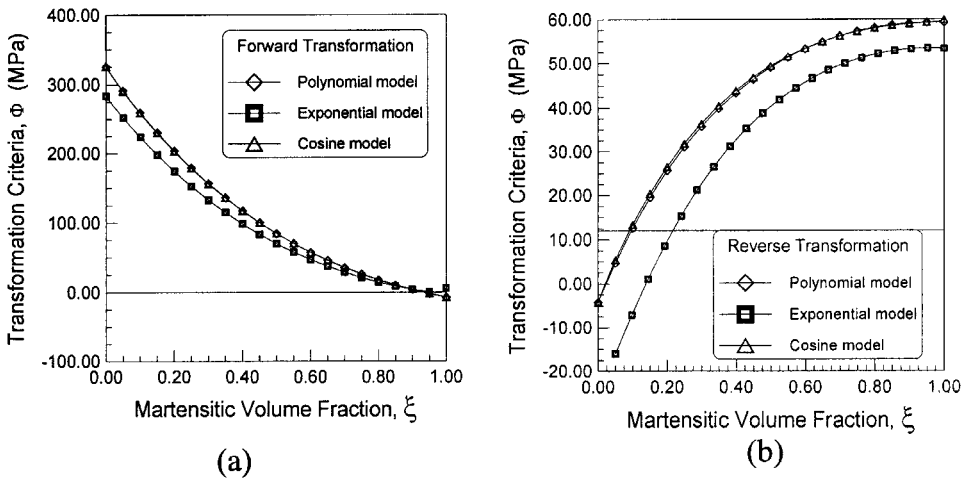


Figure 18. (a) Convergence of Φ to zero with increasing ξ in forward transformation correction for different SMA constitutive models; (b) convergence of Φ to zero with changing ξ in reverse transformation correction for different SMA constitutive models

4.2. SMA torque tube cases

In this section, BVPs are analysed where transformation may not be uniform and combined loading in the stress space takes place. The accuracy properties of the algorithms are then assessed based on the error as a function of the imposed strain increment size.

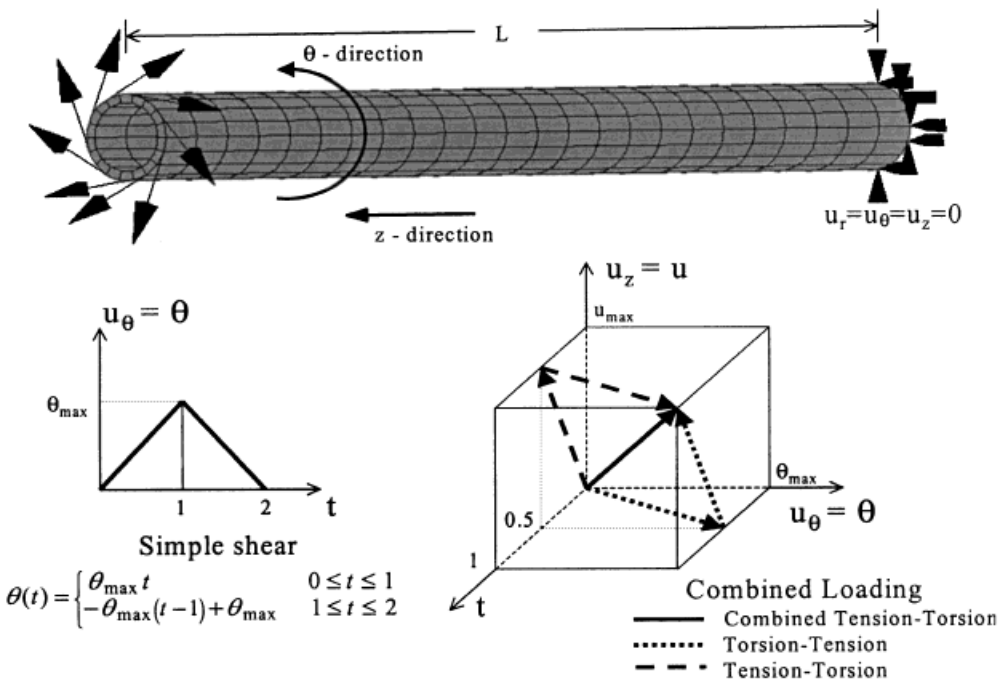


Figure 19. Schematic and loading history of the fourth boundary value problem and the loading histories of the combined loading cases

4.2.1. Torsion-induced pseudoelasticity. The schematic of the 3-D BVP and the loading history are shown in Figure 19, consisting of a SMA torque tube loaded in simple shear (torsion). The domain is modelled using 24 special eight node quadratic axisymmetric elements with an additional twist degree of freedom along the length ($L = 82$ mm).¶ Only one element is used in the radial direction, since the tube is very thin ($r_i = 2{\cdot}5$ mm, $r_o = 3{\cdot}17$ mm).‖ The tube is fully fixed at the right end and angular (rotation) boundary condition is applied at the other end. The tube is initially at the austenitic finish temperature and the maximum value of the applied rotation (1 rad/57·3°) is such that full forward transformation throughout the tube takes place. The rotations are then relaxed to zero to obtain full reverse-phase transformation. The material properties used for the SMA torque tube are given in Table I, except for the maximum transformation strain, H, which is taken to be 1 per cent. It should be realized here that since the only non-zero stress is the in-plane shear stress, $\sigma_{z\theta}$, which varies linearly in the radial direction, phase transformation will not be uniform and will also vary almost linearly radially. Concentric circles of material with the same value of ξ will form with higher values of ξ in the outer layers (see Figure 20(a)).

¶The purpose is to model a torque tube which geometrically resembles those in commercial service. The specifications used here are based on a torque tube manufactured by Memory Corp. Therefore, to obtain suitable aspect ratio, 24 elements are used in the axial direction even though there is no axial variation in stress

‖Due to the thinness of the tube in the radial direction, a very small variation of martensitic volume fraction is expected. A single quadratic element is deemed sufficient to capture the resulting variation

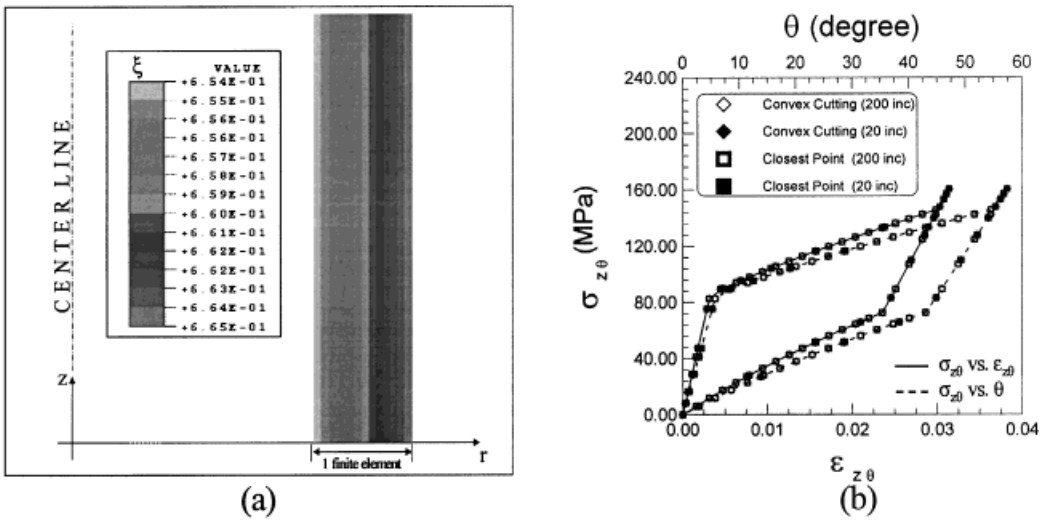


Figure 20. (a) Formation of concentric circles of the martensitic volume fraction, ξ, due to stress ($\sigma_{z\theta}$) variation in radial direction; (b) *average* in-plane shear stress versus *average* in-plane shear strain for two different increment sizes (problem 4)

In Figure 20(b), the *average*** shear stress, $\sigma_{z\theta}$, of a finite element is plotted against the *average* shear strain, $\varepsilon_{z\theta}$, and the applied rotation, θ, respectively. Only selective data points are shown on the figure for clarity of presentation. The results are plotted for both algorithms for two cases of 20 and 200 strain increments each, respectively. Defining the normalized strain increment as the ratio of the strain increment to the strain at the start of the forward transformation at $A^{\circ f}$, 20 and 200 increments correspond to 1·0 and 0·1 normalized strain increments, respectively. Both algorithms give satisfying results regardless of the increment size and 4–5 iterations were necessary for a complete transformation correction in all the cases regardless of the size of increment. Even though the polynomial model suggests linear hardening, the relationship shown is not such because transformation does not occur simultaneously in the radial direction. Note that it took approximately 54 degree of rotation†† to obtain full forward phase transformation and 1 per cent of transformation strain.

4.2.2. Torsion-tension-induced pseudoelasticity. The geometry and the finite element discretization of the BVP remains the same as shown in Figure 19, except that the fixed end is now fixed only in the axial and angular directions. The radial direction is free to move. This ensures that the axial stress (σ_{zz}) is constant and the torsional stress ($\sigma_{z\theta}$) varies only in the radial direction

**The *average* stress is obtained through interpolation at the centroid of a finite element. Effectively, averaging takes place only radially because stress distribution is uniform in the axial direction

††The large rotation results in a small strain and no rigid-body rotations are superimposed. In fact, the non-linear Green–Lagrange shear strain is equal to the engineering shear strain (3·25 per cent) in this case. This is why, the use of small-strain-based model is justified here. To take care of motions with large rotation component and small strain expected in SMA-based devices like a bending beam, can be effectively taken care of by using the same model but employing Green–Lagrange strain and second Piola–Kirchhoff stress. This has been done in a recent SPIE proceeding paper by Govindjee and Hall [48] and by the present authors in a paper that is already submitted [49]

throughout the tube length and that no radial stresses develop. Displacement (axial) and angular (rotation) boundary values are applied at the other end of the tube. The boundary displacement and rotation are applied such that at $A^{\circ f}$, the whole tube transforms from austenite to martensite, followed by unloading so that full reverse transformation takes place. The boundary conditions are applied as a function of the loading parameter, t, as shown in Figure 19 in the following three ways: (1) combined (simultaneous) tensile and torsional loading and unloading, (2) first torsional and then tensile loading, and (3) first tensile and then torsional loading. The values for maximum rotation, $\theta_{\max}$, and maximum axial displacement, $u_{\max}$, which appear in the loading history are chosen as $34{\cdot}5^{\circ}$ ($\sim 2{\cdot}3$ per cent in-plane shear strain) and $3{\cdot}5$ mm ($\sim 4{\cdot}3$ per cent axial strain), respectively. Only the loading part of the combined tension-torsion is shown in Figure 19. The results from the first case are also used to examine the performance of the algorithms, whereas the last two cases along with the first one are used to exhibit the path dependence of the material response for the forward transformation only, since similar observations can be made for the reverse transformation. The material properties given in Table I are used, except for the maximum transformation strain, H, which is taken to be 3 per cent.

The combined tensile and torsional loading case is analysed for both algorithms using 20 $(2{\cdot}0), 100(0{\cdot}4), 200(0{\cdot}2)$ and 1000 $(0{\cdot}04)$ strain increments (normalized strain increments) in order to evaluate the numerical accuracy of the algorithms. Figures 21(a) and 21(b) show the plots of average axial stress, σ_{zz} vs. average axial strain, ε_{zz}, and average shear stress, $\sigma_{z\theta}$ vs. average shear strain, $\varepsilon_{z\theta}$, respectively, for the convex cutting plane algorithm. Only selective data points are used for clarity of presentation. Similar to the previous boundary value problem, the transformation is not uniform in the radial direction due to the non-uniformity of $\sigma_{z\theta}$. The plots show that the determination of the constitutive response is independent of the increment size. The same analysis has been made for the closest point projection algorithm with similar results not shown in this paper. Figures 22(a) and 22(b) enhance this point, where the same variables are plotted for both algorithms for the lowest (20) and highest (1000) number of increments. To quantify the numerical error in the results, an error analysis is carried out in which the measure of numerical error is taken to be [42]

$$\text{error} = \frac{\|\boldsymbol{\sigma} - \boldsymbol{\sigma}^{\text{ex}}\|}{\|\boldsymbol{\sigma}^{\text{ex}}\|} \times 100 \tag{76}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^{\text{ex}}$ are the computed and exact values of the stress tensor, and $\|\cdot\|$ is the inner product of the enclosed quantity. The exact values are those which are calculated employing a sufficiently small normalized strain increment below which the values do not change. In this case the normalized strain increment is $0{\cdot}02$, corresponding to 2000 strain increments. Since the stress state is inhomogeneous in the material, an arbitrary Gauss integration point is chosen to obtain the stress values just before the end of the full transformation. The result of the error analysis is given in Figures 23(a) and 23(b) for both forward and reverse transformations, respectively. It is observed that even for the largest normalized strain increment of 2, the error is not more than $0{\cdot}004$ per cent in the forward transformation and $0{\cdot}2$ per cent in the reverse transformation for both schemes, which are negligible. Finally, in Figure 24, the rate of convergence of the return path to the transformation surface is illustrated. For both algorithms, convergence is seen to be achieved in not more than five iterations even for large strain increments. This result is true for both forward and reverse transformations.

To study the path dependence of phase transformation, all three cases mentioned above are compared for the forward-phase transformation. At the end of the loading period, full forward

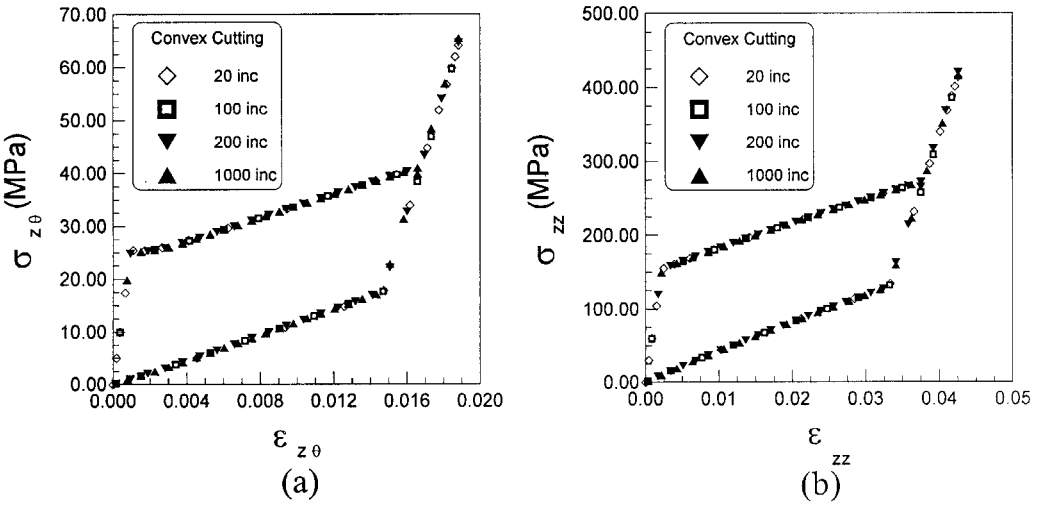


Figure 21. (a) Average shear stress versus average shear strain; (b) average axial stress versus average axial strain for four different increment sizes using the convex cutting plane algorithm—simultaneous tensile-torsional loading (problem 5)

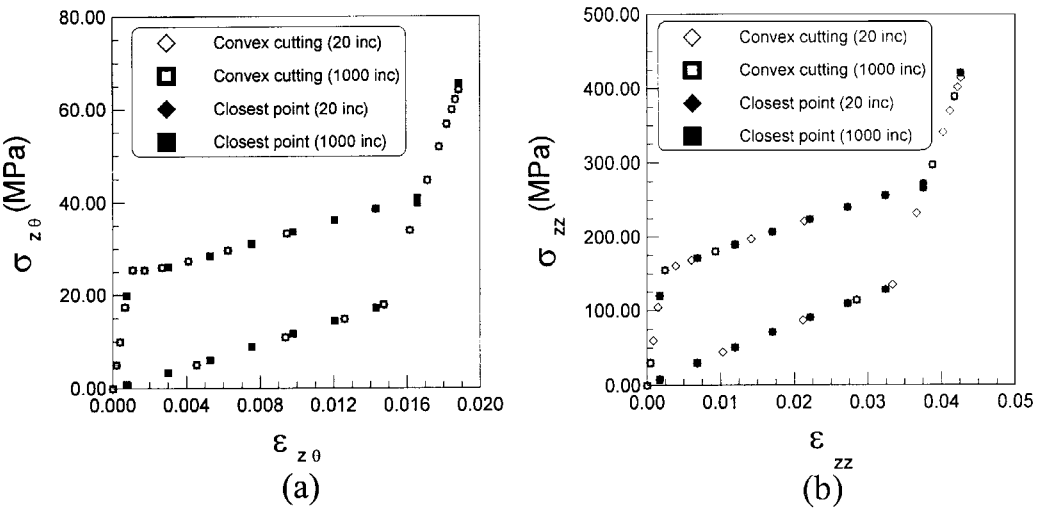


Figure 22. (a) Average shear stress versus average shear strain; (b) average axial stress versus average axial strain for the smallest and largest increment size using both integration algorithms—simultaneous tensile-torsional loading (problem 5)

transformation is achieved in all cases. In Figures 25(a), (b), and (c), $\sigma_{z\theta}$, σ_{zz} and ξ are plotted against loading parameter, t. The evolution of both stresses is most easily recognizable for the combined loading case where an elastic loading of austenite is followed by phase transformation and then an elastic loading of martensite (Figures 25(a) and (b)). In the torsion-tension case, when the rotation is applied initially, $\sigma_{z\theta}$ exhibits an elastic response followed by phase transformation

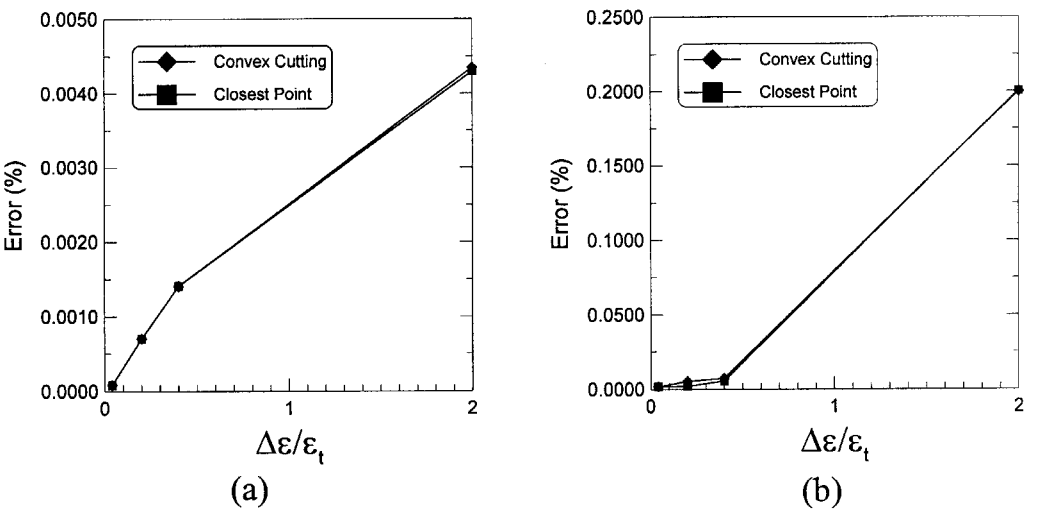


Figure 23. Per cent error versus increasing normalized strain increment sizes for both algorithms during (a) forward-phase transformation; (b) reverse-phase transformation (problem 5)

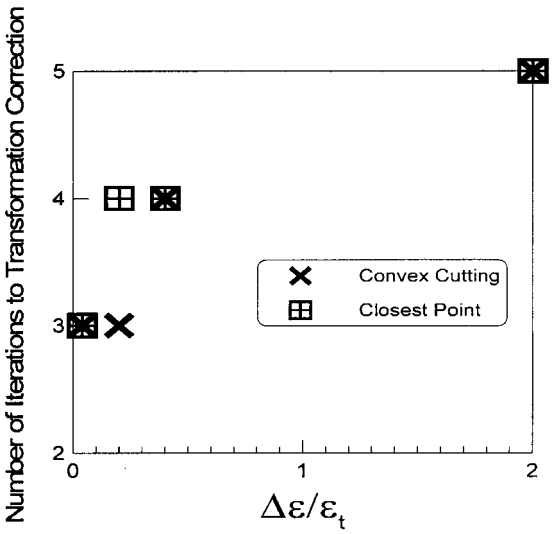


Figure 24. Number of iterations to transformation correction versus increasing normalized strain increment sizes for both algorithms (problem 5)

(Figure 25(a)) while σ_{zz} remains zero (Figure 25(b)). At $t = 0{\cdot}5$ when axial displacement is applied, a non-linear relaxation results in the rotated tube to an extent where all the torsional stress is removed from the tube even though the rotational boundary condition is still intact (Figure 25(a)). On the other hand, for tension-torsion case, $\sigma_{z\theta}$ is zero in the first half of the loading and then increases elastically due to the applied rotation (Figure 25(a)). This is because full forward transformation has already taken place during the first loading half when axial

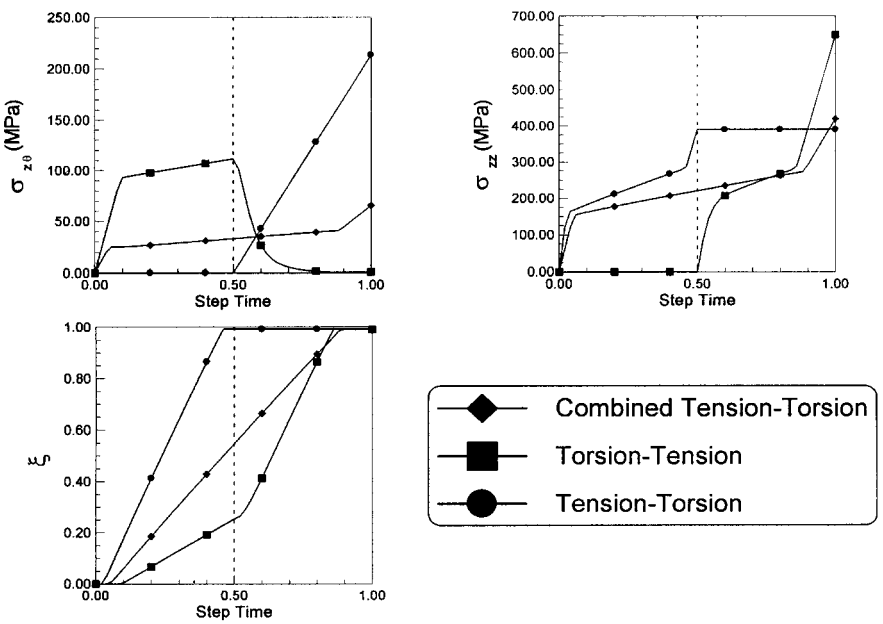


Figure 25. (a) In-plane shear stress; (b) axial stress; (c) martensitic volume fraction versus loading parameter for combined tension-torsion, torsion-tension, and tension-torsion

displacements are applied (Figure 25(c)). This is also manifested by σ_{zz}, which remains constant during the rotation after exhibiting an elastic-transformation-elastic behaviour during the applied axial displacement (Figure 25(b)).

The graph of the martensitic volume fraction, ξ, versus the loading parameter, t, in Figure 25(c) and the evolution of the four components of transformation strain ($\varepsilon^t_{z\theta}, \varepsilon^t_{zz}, \varepsilon^t_{rr}, \varepsilon^t_{\theta\theta}$) vs. the loading parameter, t, for each case in Figure 26(a), (b) and (c) also support the above-mentioned results. The remaining two components of transformation strain, ε^t_{rz} and $\varepsilon^t_{r\theta}$ are zero and they are not plotted. Note the volume conservation of transformation strain by the observation that $\varepsilon^t_{rr} = \varepsilon^t_{\theta\theta} = -0{\cdot}5\varepsilon^t_{zz}$ for all cases. In the combined tension-torsion case, all the non-zero transformation strains evolve simultaneously. In Figure 26(a), the increase in $\varepsilon^t_{z\theta}$ for the torsion-tension case justifies the relaxation of $\sigma_{z\theta}$ to zero in Figure 25(b). Also, note that there is no generation of $\varepsilon^t_{z\theta}$ in tension-torsion case because maximum transformation strain is already achieved during the axial displacement resulting in purely elastic response in the rotation part (Figure 26(a)).

Finally note that all the above-mentioned results are influenced by the choice of maximum rotation, $\theta_{\max}$, and maximum axial displacement, $u_{\max}$. For example, there is a possible selection of $\theta_{\max}$ and $u_{\max}$ for which the rate of transformation is the same for both tension and rotation.

Until now, tests have shown that both algorithms generate consistent results for all cases performed. One conclusion that can be reached by looking at Boxes 2 and 3 is that the closest point projection algorithm requires more computations than the cutting plane algorithm. This is corroborated in Table II, where the closest point algorithm is shown to consume more CPU time for each analysis, ranging approximately from 6 to 160 per cent of the CPU time consumed by the cutting plane algorithm. As mentioned in Section 3.4, the cutting plane algorithm only requires function evaluations, whereas the closest point algorithm requires the solution of a set of non-linear

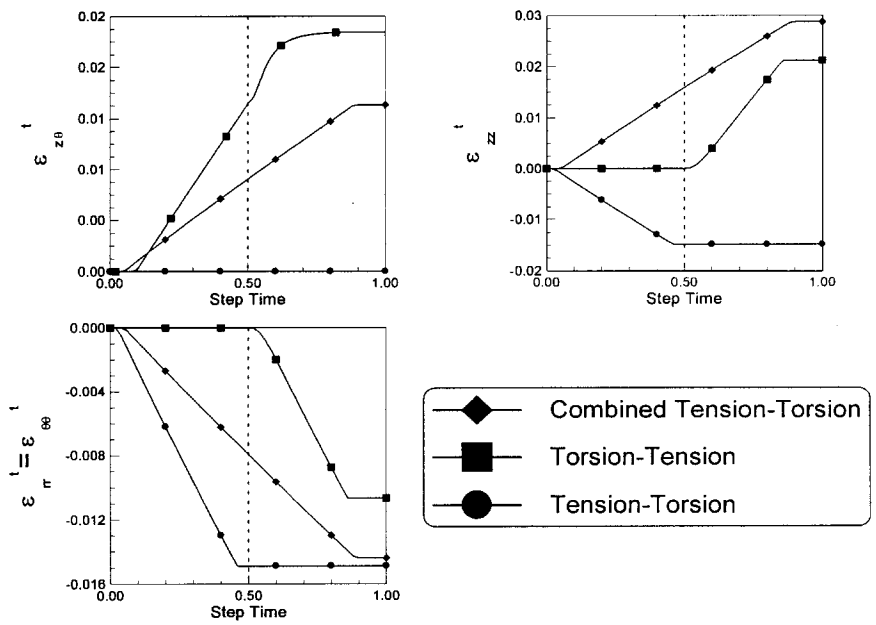


Figure 26. Various components of transformation strain versus loading parameter for combined tension-torsion, torsion-tension, and tension-torsion: (a) in-plane shear transformation strain; (b) axial transformation strain; (c) radial and tangential transformation strain

Table II. CPU time comparison for problems 1–4 for the same number of increments

		Cutting plane algorithm		Closest point algorithm	
Problem	Inc.	Glo. iter.	CPU time (s)	Glo. iter.	CPU time (s)
Isothermal uniaxial pseudoelasticity	33	41	2·70	40	3·40
Uni. const. stress therm. induced trans.	41	45	2·69	44	3·61
Uni. therm. induced trans. with a spring	13	48	3·88	46	4·10
Torsion-induced pseudoelasticity	20	22	7·52	22	10·01
	200	200	50·00	200	76·66
Torsion-tension-induced pseudoelasticity	20	29	10·07	29	15·82
	100	110	40·70	109	106·33
	200	207	61·63	206	161·63
	1000	1002	284·16	1001	464·72

SGI power challenge dual RL10000 processors are used for numerical simulation

algebraic equations, which results from the backward Euler integration of the transformation strain flow rule. The solution to this system of equations is then found using Newton's iteration method which invloves in addition to tensorial operations (multiplications, tensor products and inner products), the evaluation of the gradient of the transformation tensor (equation (60)) and the inversion of the algorithmic tangent tensor during forward transformation (equation (55)).

In the present study, LU decomposition and backsubstitution method is employed to find the inverse of the algorithmic tangent tensor because of its inherent stability due to pivoting. Due to

the specific form of equations (55) and (60), an alternative is to apply the Sherman–Morrison–Woodbury (SMW) formula to find the closed-form solution of the algorithmic tangent tensor. The advantage of employing this formula is that lesser number of operations is required to find the inverse than the direct methods, such as LU decomposition and backsubstitution, but it is less stable. An estimation of the number of operations involved in the constitutive subroutine revealed that approximately $\frac{1}{15}$th of the total CPU time would have been saved if SMW formula were employed. This observation does not adversely affect the results shown in Table IV for the CPU time cost of the finite element solution of the BVPs, therefore, LU decomposition and backsubstitution method is employed because with a minimal extra effort, stability is assured.

Table II also contains the number of increments and total number of iterations used by the global finite element (FE) program. Since the number of increments used by the FE code were fixed, the total number of iterations suggests the global convergence speed. In that regard, it is noted that almost the same global convergence behaviour is obtained for both algorithms. This implies that the global convergence rate is almost preserved even when inconsistent tangent tensor is employed for the cutting plane algorithm. This may be qualified by the observation that the numerical problems that are solved in this study are dominated more by the cost of stress evaluation rather than equation solving where one may need a proper tangent tensor which is unavailable in a cutting plane algorithm. It may be concluded then that the convex cutting plane algorithm is more suitable for solving the type of SMA-based boundary value problems studied in this paper and for the class of SMA constitutive models which show convergence behaviour during transformation correction as shown in Figures 18(a) and 18(b). However, for more complex SMA constitutive models which incorporate pressure sensitivity, softening involving transformation surfaces with corners, transformation hardening described by non-convex functions, etc., the transformation function may not be as smooth and monotonic with respect to changing martensitic volume fraction. In that case, the closest point projection algorithm will be a better choice due to its implicit nature, which results in more stability.

5. CONCLUSION

In this study, the return mapping algorithms are evaluated at length in the context of shape memory alloys. Two return mapping algorithms are studied in this work—the closest point projection algorithm and the convex cutting plane algorithm. The general transformation correction (return mapping) structure and its geometrical interpretation are presented. A summary of the polynomial thermomechanical constitutive model based on the unified thermodynamic model of Lagoudas *et al.* [12] is presented with the derivation of the continuum tangent moduli. Detailed derivation of the algorithms based on the SMA thermomechanical constitutive model is given. The consistent tangent moduli is derived for the case of closest point projection algorithm. The same is not possible for the convex cutting plane algorithm and continuum tangent moduli have to be used. In addition to the polynomial model, the cosine and exponential models have been implemented under the umbrella of the unified thermodynamic model for the convex cutting plane algorithm. It is shown that the application of the two algorithms on these SMA constitutive models will exhibit similar convergence behaviour during transformation correction (return mapping).

Five SMA boundary value problems (BVPs) are used to analyse the convergence characteristics during transformation correction and the numerical accuracy of the algorithms. The first three boundary value problems are uniaxial stress-induced pseudoelasticity, constant stress thermally

induced phase transformation and variable stress thermally induced phase transformation. They are analysed to study transformation-stress state during transformation correction, together with assessing the convergence characteristics of the iteration method employed. It is shown that both algorithms produce consistent results for the same number of increments. The salient features of the transformation correction are portrayed with the help of plotting the transformation surface in stress or stress-temperature space, during an arbitrary time increment at each Newton iteration. The convergence behaviour of the algorithms is further understood by plotting the transformation function against the martensitic volume fraction for such an increment.

Two additional BVPs based on SMA torque tubes are performed at varying number of increments to analyse the numerical accuracy of the algorithms. Separate pure torsion and combined tension-torsion displacement boundary conditions are applied on the SMA torque tube to simulate loading paths of varying complexity. It is shown that both algorithms give satisfactory results for any number of increments validating their accuracy. An analysis is then carried out to estimate the numerical error in the stress solution and they are found to be negligible for both schemes. Within the combined loading test, three more cases are simulated to analyse the path dependence of the state variables on the loading.

Overall, it is noted that the closest point projection algorithm involves more computations than the cutting plane algorithm since in addition to tensorial operations, the evaluation of the gradient of the transformation tensor in the flow rule and also inversion of algorithmic tangent tensor during forward transformation are required. This is supported by the computation cost of each BVP for both algorithms. The effect of the inconsistent tangent tensor on the global convergence rate is also not observed in the case of cutting plane algorithm. This may be the consequence of the studied BVPs, which are dominated more by stress evaluation than equation solving, where a proper tangent tensor may be required. With this in mind, it is concluded that the convex cutting plane algorithm, even though it is an explicit integration algorithm, is found to be more preferable for this class of SMA constitutive model based on the computation time it took to solve the studied BVPs numerically.

APPENDIX

I.1. Remarks on implementation of the algorithms

Since return mapping algorithms are essentially strain driven, they can be readily incorporated into the displacement based finite element method. At this point, it is assumed that the incremental SMA constitutive relations resulting from the application of the return mapping algorithms hold pointwise (or at Gauss integration points). In an incremental finite element analysis at a time $t \in [t_n, t_{n+1}]$, previously converged values of state variables $\{\boldsymbol{\varepsilon}_n, T_n, \boldsymbol{\varepsilon}^t_n, \xi_n\}$ at time $t = t_n$ are supplied to the return mapping algorithm along with the imposed increments $\{\Delta\boldsymbol{\varepsilon}_{n+1}, \Delta T_{n+1}\}$. Using the constitutive response, the remaining state variables $\boldsymbol{\varepsilon}^t_{n+1}, \xi_{n+1}$ are updated for $t = t_{n+1}$, along with the stress, $\boldsymbol{\sigma}_{n+1}$, and the tangent moduli tensors, $\mathscr{L}_{n+1}, \boldsymbol{\Theta}_{n+1}$.

As this implementation is performed, it is computationally cost-effective if the rank of the tensors appearing in the algorithms is reduced. Therefore, rank-two tensors are reduced to column vectors and rank-four tensors are mapped onto matrices, following standard practices of FEM. Care must be exercised in these contraction mappings when it comes to non-conventional tensors, like $\partial_{\boldsymbol{\sigma}}\boldsymbol{\Lambda}$ in the closest point projection algorithm.

I.1.1. Loading/unloading conditions in stress space. For imposed increments of $\{\Delta\boldsymbol{\varepsilon}_{n+1}, \Delta T_{n+1}\}$, it is necessary to identify whether the deformation is elastic or transforming. The transformation function can be used to identify the elastic and transformation regions, and also the loading and unloading criteria to determine the nature of the deformation. This requires that the direction of loading in the stress-temperature space with respect to the transformation surface be calculated. To carry out these calculations, the transformation function defined in equation (13) is written as follows:

$$\Phi(\boldsymbol{\sigma}, T, \xi) = \Psi(\boldsymbol{\sigma}, T) - \kappa(\xi) \tag{77}$$

where $\kappa(\xi)$ contains the transformation hardening terms associated with the martensitic volume fraction, ξ. It is recalled that the first step in the return mapping algorithms is to calculate the thermoelastic predictor, $\boldsymbol{\sigma}^{(0)}$. Then for time $t \in [t_n, t_{n+1}]$, we can define

$$\begin{aligned} \Psi_{n+1}^{(0)} &:= \Psi(\boldsymbol{\sigma}_{n+1}^{(0)}, T_{n+1}) = \boldsymbol{\sigma}_{n+1}^{(0)}\boldsymbol{\Lambda}_n + \tfrac{1}{2}\boldsymbol{\sigma}_{n+1}^{(0)} : \Delta\mathcal{S} : \boldsymbol{\sigma}_{n+1}^{(0)} + \boldsymbol{\sigma}_{n+1}^{(0)} : \Delta\boldsymbol{\alpha}(T_{n+1} - T_0) + \rho\Delta s_0 T_{n+1} \\ \Psi_n &:= \Psi(\boldsymbol{\sigma}_n, T_n) = \boldsymbol{\sigma}_n\boldsymbol{\Lambda}_n + \tfrac{1}{2}\boldsymbol{\sigma}_n : \Delta\mathcal{S} : \boldsymbol{\sigma}_n + \boldsymbol{\sigma}_n : \Delta\boldsymbol{\alpha}(T_n - T_0) + \rho\Delta s_0 T_n \end{aligned} \tag{78}$$

The term $\kappa(\xi)$ is neglected in this formulation because ξ remains the same for the converged solution at t_n and the thermoelastic prediction at t_{n+1}. This is true for the transformation strain also.

Designating Φ^{f} and Φ^{r} to be transformation functions (cf. equation (13)) for the forward and reverse directions, respectively: the following scheme is now used to determine the different loading conditions:

```
if Ψ^(0)_{n+1} − Ψ_n > 0          (determines if loading is taking place)
    if Φ^(0)f_{n+1} ⩽ 0           (determines if elastic or transformation loading is taking place)
        elastic
    else
        forward transformation
elseif Ψ^(0)_{n+1} − Ψ_n < 0      (determines if unloading is taking place)
    if Φ^(0)r_{n+1} ⩽ 0           (determines if elastic or transformation unloading is taking place)
        elastic
    else
        reverse transformation
elseif Ψ^(0)_{n+1} − Ψ_n = 0      (determines if neutral loading is taking place)
        elastic
endif.
```

Note that the above scheme is an implementation of the transformation consistency condition given by

$$\begin{aligned} \dot{\Phi} &= \Psi(\boldsymbol{\sigma} + d\boldsymbol{\sigma}, T + \mathrm{d}T) - \Psi(\boldsymbol{\sigma}, T) = \Psi(\boldsymbol{\sigma}_{n+1}^{(0)}, T_{n+1}) - \Psi(\boldsymbol{\sigma}_n, T_n) \\ &= \partial_{\boldsymbol{\sigma}}\Psi : \mathrm{d}\boldsymbol{\sigma} + \partial_T\Psi\,\mathrm{d}T \end{aligned} \tag{79}$$

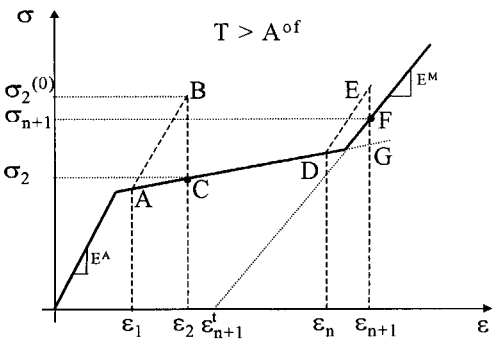


Figure 27. Stress–strain curve showing the end of transformation problem during return mapping

I.1.2. End of transformation. Another issue which is related to the above discussion and important in the context of numerical simulation is the hypothesis of convex connected elastic domains [25], which is used in proving that whenever the trial state violates the transformation consistency condition, the step is inelastic. However, in the case of shape memory alloys, the elastic domain is not connected as such, e.g. when the material is loaded above A^{of}, it initially behaves fully elastic and then forward transformation starts and is followed by another fully elastic response. In Figure 27, the thermoelastic prediction originating at point A violates the transformation consistency condition and the material transforms. The return mapping algorithm corrects the stress and brings it down to point C. However, at point D, the strain increment is such that the end of the increment at point F, the material becomes fully elastic. The application of the return mapping algorithm will result in an erroneous solution defined by point G. Two solutions can be proposed to rectify this situation, both of which have been successfully used by the authors.

In one solution, a simple forward Euler integration scheme is used to define this transition from the phase transformation region to the fully elastic behavior. A flag is introduced in the coding scheme which activates this integration routine when the phase transformation is almost near its end. For example, in the case of forward transformation during a strain increment, if the trial thermoelastic (prediction) state violates the transformation criteria and in the subsequent transformation correction $\xi \geqslant 0{\cdot}995$, the secondary forward Euler integration scheme takes over from the beginning of the strain increment. Further analysis is then carried out by subincrementing the imposed strain increment into a sufficiently large number of increments. This may result in inefficiency and loss of accuracy if the original increment size is large.

Another solution to calculate the material response for a given strain increment such as shown in Figure 27 at point D is as follows. During the return mapping if ξ becomes greater than 1·0, it is assumed that the material is fully elastic. In that case, the material constants and the transformation strain are updated to the expected values at the end of the transformation ($\xi = 1$) as shown in Figure 27 and the stress is then calculated based on the updated values. Auricchio and Sacco [25] independently proposed a similar solution to this problem for an isothermal 1-D case by introducing another trial stress associated with complete transformation. If the first trial stress violates the condition for the start of transformation and the second trial stress does not violate the condition for the end of transformation, then enforcement of the consistency condition will guarantee a computational result. Otherwise as mentioned above, the material constants and the transformation strain are updated to the expected values at the end of the transformation and the

stress is calculated. The only drawback in these schemes is their implementation in a full 3-D case, where it is generally not known *a priori* how each component of transformation strain tensor (or any internal state variable) evolves with the applied loading. This precludes one from knowing the final values of these components at the end of transformation and thus updating them consistently. Presently, the 3-D implementation is performed by updating the transformation strain-tensor-based on the transformation tensor, $\mathbf{\Lambda}$, in the last return mapping iteration before ξ became greater than 1·0.

I.2. Derivation of consistent tangent stiffness moduli for closest point projection algorithm

Taking the total differential of the constitutive stress–strain relation in equation (41) and the discrete flow rule in equation (42)

$$\mathrm{d}\boldsymbol{\sigma}_{n+1} = \mathcal{S}_{n+1}^{-1}\{\mathrm{d}\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_{n+1}\mathrm{d}T_{n+1} - \mathrm{d}\boldsymbol{\varepsilon}_{n+1}^{t} - [\Delta\mathcal{S} : \boldsymbol{\sigma}_{n+1} + \Delta\boldsymbol{\alpha}(T_{n+1} - T_0)]\mathrm{d}\xi_{n+1}\} \tag{80}$$

$$\mathrm{d}\boldsymbol{\varepsilon}_{n+1}^{t} = \mathbf{\Lambda}_{n+1}\mathrm{d}\xi_{n+1} + \begin{cases} (\xi_{n+1} - \xi_n)\partial_{\boldsymbol{\sigma}}\mathbf{\Lambda}_{n+1} : \mathrm{d}\boldsymbol{\sigma}_{n+1}, & \dot{\xi} > 0 \\ \mathbf{0}, & \dot{\xi} < 0 \end{cases} \tag{81}$$

For forward transformation, by substituting the expression for $\mathrm{d}\boldsymbol{\varepsilon}_{n+1}^{t}$ into equation (80) an algorithmic relation for $\mathrm{d}\boldsymbol{\sigma}_{n+1}$ in terms of $\mathrm{d}\boldsymbol{\varepsilon}_{n+1}$, $\mathrm{d}T_{n+1}$ and $\mathrm{d}\xi_{n+1}$ is obtained, i.e.

$$\mathrm{d}\boldsymbol{\sigma}_{n+1} = \mathscr{E}_{n+1} : [\mathrm{d}\boldsymbol{\varepsilon}_{n+1} - \boldsymbol{\alpha}_{n+1}\mathrm{d}T_{n+1} - \partial_{\boldsymbol{\sigma}}\Phi_{n+1}\mathrm{d}\xi_{n+1}] \tag{82}$$

where equation (15) is used to substitute $\partial_{\boldsymbol{\sigma}}\Phi_{n+1}$ and $\mathscr{E}_{n+1}$ is the algorithmic stiffness tensor defined as

$$\mathscr{E}_{n+1} := [\mathcal{S}_{n+1} + (\xi_{n+1} - \xi_n)\partial_{\boldsymbol{\sigma}}\mathbf{\Lambda}_{n+1}]^{-1} \tag{83}$$

On the other hand, differentiation of the discrete consistency condition yields

$$\mathrm{d}\Phi = \partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathrm{d}\boldsymbol{\sigma}_{n+1} + \partial_T\Phi_{n+1}\mathrm{d}T_{n+1} + \partial_{\xi}\Phi_{n+1}\mathrm{d}\xi_{n+1} = 0 \tag{84}$$

Now substitution of $\mathrm{d}\boldsymbol{\sigma}_{n+1}$ into the above equation results in an expression for $\mathrm{d}\xi$:

$$\mathrm{d}\xi_{n+1} = \frac{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \mathrm{d}\boldsymbol{\varepsilon}_{n+1} + (\partial_T\Phi_{n+1} - \partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \boldsymbol{\alpha}_{n+1})\mathrm{d}T_{n+1}}{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1} - \partial_{\xi}\Phi_{n+1}} \tag{85}$$

Consequently, substitution of equation (85) into equation (82) yields the expression

$$\begin{aligned} \mathrm{d}\boldsymbol{\sigma}_{n+1} = & \left[\mathscr{E}_{n+1} - \frac{\mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1} \otimes \mathscr{E}_{n+1} : \partial_{\xi}\Phi_{n+1}}{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1} - \partial_{\xi}\Phi_{n+1}}\right] : \mathrm{d}\boldsymbol{\varepsilon}_{n+1} \\ & + \mathscr{E}_{n+1} : \left[\partial_{\boldsymbol{\sigma}}\Phi_{n+1}\left(\frac{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \boldsymbol{\alpha}_{n+1} - \partial_T\Phi_{n+1}}{\partial_{\boldsymbol{\sigma}}\Phi_{n+1} : \mathscr{E}_{n+1} : \partial_{\boldsymbol{\sigma}}\Phi_{n+1} - \partial_{\xi}\Phi_{n+1}}\right) - \boldsymbol{\alpha}_{n+1}\right]\mathrm{d}T_{n+1} \end{aligned} \tag{86}$$

Defining the algorithmic tangent stiffness tensor, $\mathscr{L}$, and the tangent thermal moduli tensors, $\mathbf{\Theta}$, as

$$\mathscr{L} := \left.\frac{\mathrm{d}\boldsymbol{\sigma}}{\mathrm{d}\boldsymbol{\varepsilon}}\right|_{n+1} \qquad \mathbf{\Theta} := \left.\frac{\mathrm{d}\boldsymbol{\sigma}}{\mathrm{d}T}\right|_{n+1} \tag{87}$$

and then using (86), the expressions for $\mathscr{L}$ and $\mathbf{\Theta}$ are obtained as given in Section 3.2.4.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the Air Force Office of Scientific Research under Grant No. F49620-98-1-0041, monitored by Major Brian Sanders.

REFERENCES

1. Wayman CM. Phase transformations, nondiffusive. In *Physical Metallurgy*, Cahn RW, Haasen P (eds). North-Holland Physics Publishing: New York, 1983; 1031–1075.
2. Bo Z, Lagoudas DC. Thermomechanical modeling of polycrystalline smas under cyclic loading, Part I: theoretical derivations. *International Journal of Engineering Science* 1999; **37**:1089–1140.
3. Shaw JA, Kyriakides S. Initiation and propagation of localized deformation in elasto-plastic strips under uniaxial tension. *International Journal of Plasticity* 1998; **13**(4):837–871.
4. Miyazaki S, Imai T, Igo Y, Otsuka K. Effect of cyclic deformation on the pseudoelasticity characteristics of Ti-Ni alloys. *Metallurgical Transactions* 1997; **42**(7):115–120.
5. Sachdeva RC, Miyazaki S. Superelastic Ni-Ti alloys in orthodontics. In *Engineering Aspects of Shape Memory Alloys*, Duerig TW, Melton KN, Stökel D, Wayman CM (eds), 1990; 452–469.
6. Melzer A, Stökel D. Performance improvement of surgical instrumentation through the use of Ni-Ti materials. In *Proceedings of the 1st International Conference on Shape Memory and Superelastic Technologies*, Pelton AR, Hodgson D, Duerig T (eds). 1994; 401–409.
7. Birman V. Review of mechanics of shape memory alloy structures. *Applied Mechanics Reviews* 1997; **50**(11): 629–645.
8. Jardine AP, Kudva JN, Martin C, Appa K. Shape memory alloy Ti-Ni actuators for twist control of smart wing designs. In *Mathematics and Controls in Smart Structures*. SPIE, 1996; **2717**:160–165.
9. Liang C, Davidson F, Schetky LM, Straub FK. Applications of torsional shape memory alloy actuators for active rotor blade control-opportunities and limitations. In *Mathematics and Controls in Smart Structures*. SPIE, 1996; 91–100.
10. Rediniotis OK, Lagoudas DC, Garner L, Wilson N. Experiments and analysis of an active hydrofoil with sma actuators. *AIAA Paper No. 98-0102*, *36th AIAA Aerospace Sciences Meeting*, Reno, Nevada, 1998.
11. Lagoudas DC, Boyd JG, Bo Z. Micromechanics of active composites with sma fibers. ASME *Journal of Material Science and Technology* 1994; **116**(3):337–347.
12. Lagoudas DC, Bo Z, Qidwai MA. A unified thermodynamic constitutive model for sma and finite element analysis of active metal matrix composite. *Mechanics of Composite Materials and Structures* 1996; **3**:153–179.
13. Lagoudas DC, Moorthi D, Qidwai MA, Reddy JN. Modeling of the thermomechanical response of active laminates with sma strips using the layerwise finite element method. *Journal of Intelligent Material Systems and Structures* 1997; **8**(6):476–488.
14. Adler PH, Yu W, Pelton AR, Zadno R, Duerig TW, Barresi R. On the tensile and torsional properties of pseudoleastic ni-ti. *Scripta Metallurgica et Materialia* 1990; **24**:943–947.
15. Melton KN. Ni-ti based shape memory alloys. In *Engineering Aspects of Shape Memory Alloys*, Duerig TW, Melton KN, Stökel D, Wayman CM (eds), 1990; 21–35.
16. Lim JT, McDowell DL. Mechanical behavior of a Ni-Ti shape memory alloy under axial-torsional proportional and nonproportional loading. *Journal of Engineering Materials and Technology* 1999; **121**:9–18.
17. Tanaka K. A thermomechanical sketch of shape memory effect: one dimensional tensile behavior. *Res Mechanica* 1986; **18**:251–263.
18. Tanaka K, Nishimura F, Hayashi T, Tobushi H, Lexcellent C. Phenomenological analysis on subloops and cyclic behavior in shape memory alloys under mechanical and/or thermal loads. *Mechanica of Materials* 1995; **19**:281–292.
19. Liang C, Rogers CA. The multi-dimensional constitutive relations of shape memory alloys. *Journal of Engineering Mathematics* 1992; **26**:429–443.
20. Boyd JG, Lagoudas DC. A thermodynamic constitutive model for the shape memory alloy materials. Part I. the monolithic shape memory alloy. *International Journal of Plasticity* 1996; **12**:805–842.
21. Brinson LC, Lammering R. Finite-element analysis of the behavior of shape memory alloys and their applications. *International Journal of Solids and Structures* 1993; **30**:3261–3280.
22. Fischer D, Sun Q.-P, Tanaka K. Transformation-induced plasticity (TRIP). *Applied Mechanics Review* 1996; **49**(6): 317–364.
23. Trochu F, Qian YY. Nonlinear finite element simulation of superelastic shape memory alloy parts. *Computers and Structures* 1997; **67**:799–810.
24. Auricchio F, Taylor RL, Lubliner J. Shape-memory alloys: macromodelling and numerical simulations of the superelastic behavior. *Computer Methods in Applied Mechanics and Engineering* 1997; **146**:281–312.
25. Auricchio F, Sacco E. A one-dimensional model for superleastic shape-memory alloys with different elastic properties between austenite and martensite. *International Journal of Non-Linear Mechanics* 1997; **32**(6):1101–1114.

26. Auricchio F, Sacco E. A superelastic shape-memory-alloy beam model. *Journal of Intelligent Material Systems and Structures* 1997; **8**(6):489–501.
27. Panahandeh M, Kasper E. Coupled thermomechanical simulation of shape memory alloys. In *Mathematics and Controls in Smart Structures*, SPIE, 1997; **3039**:468–480.
28. Govindjee S, Kasper EP. A shape memory alloy model for Uranium-Niobium accounting for plasticity. *Journal of Intelligent Material Systems and Structures* 1998; **8**:815–823.
29. Reisner G, Werner EA, Fischer FD. Micromechanical modelling of martensitic transformation in random microstructures. *International Journal of Solids and Structures* 1998; **19**:2457–2473.
30. Simo JC, Ortiz M. A unified approach to finite deformation elastoplastic analysis based on the use of hyperelastic constitutive equations. *Computer Methods in Applied Mechanics and Engineering* 1985; **49**:221–245.
31. Sottos NR, Kline GE, Qidwai MA, Lagoudas DC. Analysis of phase transformation fronts in embedded shape memory alloy composites. In *Mathematics and Controls in Smart Structures*, vol. 2715, SPIE, 1996; 25–29.
32. Jonnalagadda KD, Sottos NR, Qidwai MA, Lagoudas DC. Transformation of embedded shape memory alloy ribbons. In *Mathematics and Controls in Smart Structures*, SPIE, 1997; **3039**:481–496.
33. Jonnalagadda KD, Sottos NR, Qidwai MA, Lagoudas DC. Transformation of embedded shape memory alloy ribbons. *Journal of Intelligent Material Systems and Structures* 1999; **9**:379–390.
34. Moss WC. On the computational significance of the strain space formulation of plasticity theory. *International Journal for Numerical Methods in Engineering* 1984; **20**:1703–1709.
35. Wilkins ML. Calculation of elastic-plastic flow. In *Methods of Computational Physics 3*, Alder B. *et al.* (eds). Academic Press: New York, 1964.
36. Rice JR, Tracey DM. Computational fracture mechanics. In *Proceedings of the Symposium on Numerical Methods in Structural Mechanics*, Fernes SJ (ed.). Academic Press: New York, 1973.
37. Simo JC, Hughes TJR. *Computational Inelasticity*. Springer: New York, 1998.
38. Ortiz M, Popov EP. Accuracy and stability of integration algorithms for elastoplastic constitutive relations. *International Journal for Numerical Methods in Engineering* 1985; **21**:1561–1576.
39. Argyris JH, Doltsinis JS, Knudson WC, Vaz LE, William KJ. Numerical solution of transient nonlinear problems. *Computer Methods in Applied Mechanics and Engineering* 1979; **17**/**18**:341–409.
40. Simo JC, Govindjee S. Non-linear b-stability and symmetry preserving return mapping algorithms for plasticity and viscoplasticity. *International Journal for Numerical Methods in Engineering* 1991; **31**:151–176.
41. Ortiz M, Pinsky PM. Global analysis methods for the solution of elastoplastic and viscoplastic dynamic problems. *Technical Report UCB/SESM 81/08*, Department of Civil Engineering, University of California at Berkeley, 1981.
42. Ortiz M, Simo JC. An analysis of a new class of integration algorithms for elastoplastic constitutive relations. *International Journal for Numerical Methods in Engineering* 1986; **23**:353–366.
43. Simo JC, Taylor RL. Consistent tangent operators for rate-independent elastoplasticity. *Computer Methods in Applied Mechanics and Engineering* 1985; **48**:101–118.
44. Truesdell C, Noll W. *The Non-linear Field Theories of Mechanics.* Springer: Berlin, 1965.
45. Gillet Y, Patoor E, Berveiller M. Calculation of pseudoelastic elements using a non symmetrical thermomechanical transformation criterion and associated rule. *Journal of Intelligent Materials and Technology* 1998; **9**:366–378.
46. Simo JC, Hughes TJ. General return mapping algorithms for rate-independent plasticity. In *Constitutive Laws for Engineering Materials: Theory and Applications*, Desai CS (ed.). Elsevier Science Publishing Co: Amsterdam, 1987; 221–231.
47. Lubliner J. *Plasticity Theory*. Macmillan Publishing Company: New York, 1990.
48. Govindjee S, Hall GJ. Computational aspects of solid–solid phase transformation modeling with a gibbs function. In *Mathematics and Controls in Smart Structures*, SPIE, 1999; 302–313.
49. Qidwai MA, Lagoudas DC. On thermomechanics and transformation surfaces of polycrystalline shape memory alloy materials. *International Journal of Plasticity* 1999, accepted.

